

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:20:16 ; Search time 25 seconds

(without alignments)
944.296 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626
Sequence: 1 MSKPLIKITLLICALSALML.....NRGGLSADDAKAKKTKPN 502

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386	52.8	392	1 ANIA_NEIGO	002219 neisseria g
2	468	17.8	374	1 NIR_RHOSH	053239 rhodobacter
3	461.5	17.6	378	1 NIR_ACHCY	P25006 achromobact.
4	453	17.3	379	1 NIR_RHIGA	001537 rhizobium g
5	451	17.2	363	1 NIR_PSECL	006006 pseudomonas
6	439.5	16.7	376	1 NIR_RHIME	092229 rhizobium m
7	437	16.6	377	1 NIR_RHIME	060214 rhizobium h
8	427.5	16.3	376	1 NIR_ALCPA	P81445 alcaligenes
9	424	16.1	330	1 NIR_ALCPX	P37064 escherichia
10	203.5	7.7	605	1 PCOA_ECOLI	047452 escherichia
11	170	6.5	552	1 ASO_CUCPM	P37064 cucurbita p
12	167	6.4	579	1 ASO_CUCMA	P24792 cucurbita m
13	164.5	6.3	609	1 CPAL_PSESM	P12374 pseudomonas
14	163	6.2	478	1 CYCA_GLUOX	047945 gluconobact
15	162	6.2	578	1 ASO_TOBAC	P04058 nicotiana t
16	161	6.1	148	1 C552_THETH	P04164 thermus the
17	157.5	6.0	589	1 CPAA_PSESM	P59571 pseudomonas
18	153.5	5.8	587	1 ASO_CUCSA	P14133 cucumis sat
19	146	5.6	468	1 CYCA_ACEPO	003318 acetobacter
20	145	5.5	555	1 ASO_BRANA	000624 brassica na
21	143.5	5.5	221.1	1 FAS_BOVIN	Q02007 bos taurus
22	137	5.2	2258	1 FAS_PIG	099191 sus scrofa
23	133.5	5.1	576	1 LACI_THACU	P56193 thatephor
24	131	5.0	531	1 LAC4_THACU	002081 thatephor
25	128.5	4.9	2224	1 FAS_HUMAN	P12329 homo sapien
26	124	4.7	533	1 CUBO_YERPE	082600 yerinia pe
27	124	4.7	572	1 LAC2_THACU	P02079 thatephor
28	118	4.4	520	1 FETS_YEAST	043561 saccharomyc
29	116	4.4	520	1 LAC1_AGABI	012541 agaricus bi
30	116	4.4	652	1 NOS2_PARDE	051705 parococcus
31	111.5	4.2	554	1 NTP2_TOBAC	P29162 nicotiana t
32	111	4.2	527	1 LAC3_TRAVI	099056 trameles vi
33	111	4.2	1101	1 GUNC_CELFI	P14090 cellulomona

34	109	4.2	108	1 C555_CHLIT	08x93 chlorobium
35	109	4.2	1062	1 CERU_MOUSE	Q61147 mus musculu
36	108	4.1	527	1 LACS_TRAVE	Q12717 trameles ve
37	107.5	4.1	520	1 LAC2_AGABI	012542 agaricus bi
38	107.5	4.1	1257	1 CCAA_BACTU	045754 bacillus th
39	107.5	4.1	2647	1 FLNA_HUMAN	P21333 homo sapien
40	107	4.1	1654	1 OMPB_RICRI	Q53047 r outer mem
41	106.5	4.1	485	1 IMDH_PYRAB	Q9UY49 pyrococcus
42	106.5	4.1	599	1 LAC2_THACU	002075 thatephor
43	106.5	4.1	642	1 PHSA_STRAT	051692 streptomyce
44	105.5	4.0	739	1 PUBLI_CAUCR	Q945f0 caulobacter
45	105	4.0	86	1 C555_CHLIT	P00123 chlorobium

ALIGNMENTS

RESULT 1
ID ANIA_NEIGO STANDARD; PRT; 392 AA.
AC 002219;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Major outer membrane protein Pan 1 precursor.
OS ANIA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=RI0;
RX MEDLINE=93014187; PubMed=1383156;
RA Hoehn G.T., Clark V.L.;
RT "Isolation and nucleotide sequence of the gene (ania) encoding the
RT major anaerobically induced outer membrane protein of Neisseria
RT gonorrhoeae.";
RL Infect. Immun. 60:4695-4703 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33084 / F62;
RX MEDLINE=93014188; PubMed=1398981;
RA Hoehn G.T., Clark V.L.;
RT "The major anaerobically induced outer membrane protein of Neisseria
RT gonorrhoeae, Pan 1, is a lipoprotein.";
RL Infect. Immun. 60:4704-4708 (1992).
RN [3]
RP SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
RP anchor (PROBABLE).
CC -1- SIMILARITY: Contains 2 placitocyanin-like domains.
CC -1- INDUCTION: BY ANAEROBIOSIS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

EMBL: M97926; AAA25462.1; -
DR PIR: A49208; A49208.
DR PDB: 1KXV; 27-FEB-02.
DR PDB: 1KXW; 27-FEB-02.
DR InterPro: IPR00117; Cu-oxidase.
DR InterPro: IPR001287; CuNO2_reductase.
DR Pfam: PF00394; Cu-oxidase; 2.
DR PRINTS: PR00695; CUNO2RDTASE.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Repeat; Outer membrane; Lipoprotein; Signal; 3D-structure.
FT SIGNAL 1 18
FT CHAIN 19 392 MAJOR OUTER MEMBRANE PROTEIN PAN 1.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).
FT

FT DOMAIN 101 195 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 245 346 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 368 387 4 X 5 AA TANDEM REPEATS OF A-A-S-A-P.
 FT REPEAT 368 372 1.
 FT REPEAT 373 377 2.
 FT REPEAT 378 382 3.
 FT REPEAT 383 387 4.
 SO SEQUENCE 392 AA; 40954 MW; A4707CGB923C97 CRC64;

Query Match 52.8%; Score 1386; DB 1; Length 392;
 Best Local Similarity 69.5%; Pred. No. 4,4e-90;
 Matches 264; Conservative 44; Mismatches 66; Indels 6; Gaps 3;

11 LICALALMLSGSNQADKAPKSSIVDAAKTA-NADVAASOEHOGEPLVDALVTA 69
 9 MIALSLPALAACG---GEQAAQAPAEPTPAASASASAAQATRETAGELPVIDATVTA 64
 70 PEVPPVDRDHPAKVVKETVEKVRRLADGVEYQFTFGGVPGQMIRREGDTIEVOP 129
 65 PEVPPAIDRDYPAKVVKETVEKTKMDGVEYRWTFPGDVPGMIRREGDTIEVER 124
 130 SNHPDSKMPHNVDPHAATGPGGAASFTAPGHTSTFSFKALQPLGYTHCAVAPVGMH 189
 125 SNHPSTVPHNVDPHAATGPGGAATFTAPGRTSTFSFKALQPLGYTHCAVAPVGMH 184
 190 ANGMVGLIIVEPEKGLPKVDKEYVYVMOGDFTYTKGKGEOGLQPPDEKAIREDAEVYFN 249
 185 ANGMVGLIIVEPEKGLPKVDKEYYIQGDDPTTGKKGAGLOLFPDMDKAAVEPEVYFN 244
 250 GSVGALTGENALAKVGETVRLFVNGGPNLTSSPHVIGIFDKVHFEKGGKGNHNIQTT 309
 245 GHVGSJAGDNALAKVGETVRLVYVNGGPNLTSSPHVIGIFDKVHFEKGGKGNHNIQTT 304
 310 LIPAGGAATTEFVNDVPGDVVLDHAIFFAPKNGALGILKEEENHEIYSHQOTDAVL 369
 305 IYPAGGSAIVEFVNDVPGDVVLDHAIFFAPKNGALGILKEEENHEIYSHQOTDAVL 364
 370 PEGAPQAIPTQGAQKTPAPA 389
 365 GSGAASH-PAASAPMAASAPA 383

RESULT 2
 NIR_RHOSH STANDARD; PRT; 374 AA.
 AC 053239;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
 GN NIRK.
 OS Rhodobacter alphaeroides (Rhodospseudomonas alphaeroides).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Rhodobacter.
 OX NCBI_TaxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2.4.3;
 RX MEDLINE=97175533; PubMed=9023186;
 RA Tosques I.E., Kwiatkowski A.V., Shi J., Shapleigh J.P.;
 RT "Characterization and regulation of the gene encoding nitrite
 reductase in Rhodobacter alphaeroides 2.4.3.";
 RL J. Bacteriol. 179:1090-1095(1997).
 CC -1- CATALYTIC ACTIVITY: Nitrite oxide + H(2)O + ferricytochrome c =
 nitrite + ferrocyclochrome c.
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
 II COPPER; FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
 WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
 OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
 CC PSEUDOMAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
 VITRO (BY SIMILARITY).
 CC -1- PATHWAY: Nitrate assimilation (denitrification).
 CC -1- SUBUNIT: Homotrimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
 CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
 CC ELECTRON TRANSFER FROM PSEUDOMAZURIN TO THE TYPE II COPPER SITE
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
 CC REDUCTION OF NITRITE.
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U62291; AAB05767.1; -
 DR HSSP: P25006; INIP.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR001287; CuNO2_reductase.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR PRINTS: PR00695; CUNO2REDTASE.
 KM Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
 KM Nitrate assimilation; Repeat; Periplasmic; Signal.
 FT CHAIN 1
 FT SIGNAL 38
 FT DOMAIN 39 374
 FT DOMAIN 93 189
 FT DOMAIN 254 355
 FT METAL 126 126
 FT METAL 131 131
 FT METAL 166 166
 FT METAL 167 167
 FT METAL 177 177
 FT METAL 182 182
 FT METAL 338 338
 SQ SEQUENCE 374 AA; 40308 MW; 3406B58E7DD9934 CRC64;

Query Match 17.8%; Score 468; DB 1; Length 374;
 Best Local Similarity 33.8%; Pred. No. 1.2e-25;
 Matches 133; Conservative 52; Mismatches 126; Indels 80; Gaps 15;

QY 34 KSTVDAAN-----KTANADNAAOEHQELPVIDAIVHAP-----VP 73
 DB 5 RAALVGAAALASAPLVIRTAGAQAAPQ-----LASAAPVDLSNLPVKHTLVP 53
 QY 74 PVVDRD-----PAKVVKMETVEKVRRLADGVEYQFTFGGVPGQMIRREGDTIE 126
 DB 54 PEPANAEQVAAASGVYINFEKRIIEKEVQLEDAVLQMTDGSIPGLMIVHESDYVE 113
 QY 127 VQSNHPDSKMPHNVDPHAATGPGGAASFTAPGHTSTFSFKALQPLGYTHCAVAPVG 186
 DB 114 LTLINPENTWPHNIDFHAATGALGGGLTLINPGBKVLRFKATRAAFVYHC-APGG 171
 QY 187 ----MHANGATGLIIVEKEGLP-----KVDKEYVYMOGDFTY-----TKGKYGEGSQ 231
 DB 172 PMIPMHVVSQMGAGINVLPRDGLKHGKPVYVDVYVYIGESDHYIPKDEGTY-----M 226
 QY 232 PFDMEKAIRE-----AEVYVNGSVGALTGENALAKVGETVRLFVNGGPNLTSS 282
 DB 227 RSTSTPEGYEDVAVVADTLIPSHIVFNGAVGLTSGAKAKAYGDV-LFV-HSQPKRS 284
 QY 283 SFHVIGEIFDKVHFEKGGKGN--NIIQTLLIPAGGAATTEKRVDPGVYVLDHAIFFA 339
 DB 285 RHHLIGCHDVL-WEIGKFNHAPENDLETWFRGGTAGAALVKFLQPGVAYAVNHLISA 343
 QY 340 FNKGALGILKVEEENHEIYSHKQTDVAVLPRG 372
 DB 344 VHGATASVAVLVEGEMDNL-----MEQVMPVPG 371

RESULT 3
 NIR_ACHCY STANDARD; PRT; 378 AA.
 ID NIR_ACHCY
 AC P25006;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).

CN NMR.

OS Achromobacter cycloclastes.

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Achromobacter.

OK NCBI_TaxID=223;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IAM 1013;

RX MEDLINE=96193667; PubMed=8605003;

RA Chen J.-Y., Chang W.-C., Chang T., Chang W.-C., Liu M.-Y., Payne W.J.,
le Gall J.;

RA "Cloning, characterization, and expression of the nitric oxide-
generating nitrite reductase and of the blue copper protein genes of
Achromobacter cycloclastes.";

RT Biochem. Biophys. Res. Commun. 219:423-428 (1996).

RL [2]

RN SEQUENCE OF 39-378.

RP STRAIN=IAM 1013;

RC MEDLINE=91308101; PubMed=1830217;

RX Penderson F.F., Kumar S., Adman E.T., Liu M.-Y., Payne W.J.,
le Gall J.;

RA "Amino acid sequence of nitrite reductase: a copper protein from
Achromobacter cycloclastes.";

RT Biochemistry 30:7180-7185 (1991).

RL [3]

RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).

RP MEDLINE=91320115; PubMed=1862344;

RX Godden J.W., Turley S., Teller D.C., Adman E.T., Liu M.-Y.,
Payne W.J., le Gall J.;

RA "The 2.3-A X-ray structure of nitrite reductase from Achromobacter
cycloclastes.";

RT Science 253:438-442 (1991).

RL [4]

RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

RP MEDLINE=96070866; PubMed=749203;

RX Adman E.T., Godden J.W., Turley S.;

RA "The structure of copper-nitrite reductase from Achromobacter
cycloclastes at five pH values, with NO₂-bound and with type II
copper depleted.";

RT J. Biol. Chem. 270:27458-27474 (1995).

CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =
nitrite + ferrocyclochrome c.

CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
II COPPER, FAD.

CC -1- PATHWAY: Nitrate assimilation (denitrification).

CC -1- SUBUNIT: Homotrimer.

CC -1- SUBCELLULAR LOCATION: Periplasmic.

CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
ELECTRON TRANSFER FROM PSEUDOUZURIN TO THE TYPE II COPPER SITE
OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
REDUCTION OF NITRITE.

CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----

CC EMBL; Z48635; CA88564.1; -

CC PIR; JC4648; JC4648.

DR PDB; 2NRD; 07-DEC-95.

DR PDB; 1N1A; 07-DEC-95.

DR PDB; 1N1B; 07-DEC-95.

DR PDB; 1N1C; 07-DEC-95.

DR PDB; 1N1D; 07-DEC-95.

DR PDB; 1N1E; 07-DEC-95.

DR PDB; 1N1F; 07-DEC-95.

DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR001287; CuNO₂ reductase.

DR InterPro; IPR006311; Tat.

DR Pfam; PF00394; Cu-oxidase; 2.

DR PRINTS; PR00695; CUNO2RPTASE.

DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.

KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
Nir; Nitrate assimilation; Repeat; Periplasmic; Signal; 3D-structure.

FT SIGNAL 1 38

FT CHAIN 39 378

FT DOMAIN 39 213

FT METAL 214 378

FT METAL 133 133

FT METAL 138 138

FT METAL 173 173

FT METAL 174 174

FT METAL 183 183

FT METAL 188 188

FT METAL 344 344

FT METAL 47 49

FT METAL 52 54

FT STRAND 60 61

FT STRAND 69 69

FT STRAND 76 76

FT STRAND 92 93

FT STRAND 102 102

FT STRAND 103 104

FT STRAND 105 105

FT STRAND 110 114

FT STRAND 115 116

FT STRAND 127 128

FT STRAND 133 133

FT STRAND 136 137

FT STRAND 138 139

FT STRAND 143 149

FT STRAND 152 152

FT STRAND 154 155

FT STRAND 156 163

FT STRAND 168 173

FT STRAND 177 178

FT STRAND 180 184

FT STRAND 185 187

FT STRAND 189 195

FT STRAND 197 197

FT STRAND 198 198

FT STRAND 200 201

FT STRAND 203 204

FT STRAND 207 208

FT STRAND 212 221

FT STRAND 225 225

FT STRAND 227 228

FT STRAND 231 231

FT STRAND 237 248

FT STRAND 249 250

FT STRAND 255 258

FT STRAND 259 260

FT STRAND 261 261

FT STRAND 262 265

FT STRAND 267 269

FT STRAND 271 274

FT STRAND 275 276

FT STRAND 278 285

FT STRAND 290 290

FT STRAND 292 295

FT STRAND 296 296

FT STRAND 299 303

FT STRAND 304 305

FT STRAND 308 309

FT STRAND 313 316

FT STRAND 321 321

FT STRAND 323 324

FT STRAND 324 324

FT STRAND 325 332
 FT STRAND 337 343
 FT HELIX 346 350
 FT TURN 351 352
 FT STRAND 355 361
 FT TURN 366 368
 SQ SEQUENCE 376 AA; 40771 MW; A70B52B814090EAS CRC64;

Query Match 17.6%; Score 461.5; DB 1; Length 378;
 Best Local Similarity 36.3%; Pred. No. 3.3e-25;
 Matches 134; Conservative 45; Mismatches 149; Indels 41; Gaps 12;

19 MUGSGNQDAKQAPKSTVDAAKTANADNAAQGHQELPYDIAIVTHAEVPPVVR 78
 12 MLAGALAG--AVAPLHTAQAHAAGAAAAGAAPDITLBRVKKDL-----VAPRPH 64
 79 DH-----PAKVVMETVEKVMRL-ADGVQYFWTFGGQVPGQIMIRREGDTTEVPS 130
 65 AHDOYAKTGRVVEFMTIEKKLVLDREGTEIHAMTFNGSVPGPLMVHENDYVELRLI 124
 131 NHPDSKMPNVDFAATGPGGAASFTAPGHTSTSFALOPGLVYHCAVAPVGM--- 187
 125 NEDVTLLHNIIDFHAATGALGGALTVQVNPGETTLRFATKRGVGVHC--APEGMVPM 182
 188 HANGVYGLIVEPKKGLP-----KVDKEYVVMQDFFYTK-----GKYGEQGLQF 233
 183 HTSGNGAIWLPFRDGLKDEKQPLTYDKIYVGSODPYVPDEKNGYKKTETPEAYE 242
 234 DMEKAIRE-DAEYVNGSVGALTGENALKAKVGETVRLFEVNGGPNLTSSFFHVEIFD 292
 243 DAVKAMRTLTPIHIVNGAVGALTGDHALTAAGE--RVLVHSAQNRDTRPHLIGHD 300
 293 KYHFEQKGENH--NIOTLLPAGGAATTEKVDVPGDYLVDAIFAAPKAGALIK 349
 301 YV-WATGKFRNPEDLDQETWLTIPGTAGAAFYTERQPGVAYVNNLIEAFELGAAGHK 359
 350 VEGEENHEI 358
 360 VTGEWMDL 368

RESULT 4
 NIR_RHIGA STANDARD; PRT; 379 AA.

AC 001537;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
 GN NIRU.
 OS Rhizobium galegae.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 NC NCB1_TaxID=399;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-179;
 RA MEDLINE=93175864; PubMed=8439151;
 RX Ye R.W., Fries M.R., Bezborodnikov S.G., Averill B.A., Tiedje J.M.;
 RT "Characterization of the structural gene encoding a copper-containing
 RT nitrite reductase and homology of this gene to DNA of other
 RT denitrifiers." Microbiol. 59:250-254(1993).
 RL Appl. Environ. Microbiol. 59:250-254(1993).
 CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferriochrome c =
 CC nitrite + ferriochrome c.
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
 CC II COPPER. FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER.
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
 CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
 CC PSEUDOURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
 CC VITRO (BY SIMILARITY).
 CC -1- PATHWAY: Nitrate assimilation (denitrification).
 CC -1- SUBUNIT: Homotrimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
 CC ELECTRON TRANSFER FROM PSEUDOURIN TO THE TYPE II COPPER SITE
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
 CC REDUCTION OF NITRITE.
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-12 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL; M97294; AAC79132.1; -;
 DR PIR; A48936; A48936.
 DR HSSP; P25006; NIF.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR001287; CUNO2_reductase.
 DR InterPro; IPR006311; Tat.
 DR Pfam; PF00394; Cu-oxidase; 2.
 DR PRINTS; PR00695; CUNO2RDTASE.
 DR TIGRFAMs; TIGR01409; Tat signal seq; 1.
 DR Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
 KW Nitrate assimilation; Repeat; Periplasmic; Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 379
 FT DOMAIN 33 214
 FT DOMAIN 215 379
 FT METAL 134 134
 FT METAL 139 139
 FT METAL 174 174
 FT METAL 175 175
 FT METAL 184 184
 FT METAL 189 189
 FT METAL 345 345
 SQ SEQUENCE 379 AA; 40694 MW; 090A3CBF66262F CRC64;

Query Match 17.3%; Score 453; DB 1; Length 379;
 Best Local Similarity 37.3%; Pred. No. 1.3e-24;
 Matches 118; Conservative 45; Mismatches 119; Indels 34; Gaps 10;

QY 72 VRRPVDRDH-----PAKVVMETVEKVMRLAD-GVEYQFWTFGGQVPGQIMIRREGD 123
 DB 59 VRRPVNAHTQKAESEPKVVERKMTTQEKIVDDKGTVEHAMTFDGSVPGPMIVHQDD 118
 QY 124 TIEVQFSNHPDSKMPNVDFAATGPGGAASFTAPGHTSTSFALOPGLVYHCAVA 183
 DB 119 YVELTLVNDPNTNELQHNIDFHSATGALGGALTVNPGDTAVLRFRATKAGVGVHC--A 176
 QY 184 PVGM--HANGVYGLIVEPKKGLP-----VDKEYVVMQDFFYTK-----GKYG 226
 DB 177 PAGMVPWHTSGMNGAIWLPFRDGLKDHGHELVYDKVYVVEODFYVPKDENGFKKYE 236
 QY 227 EGGLOPDMEKAIRE-DAEYVNGSVGALTGENALKAKVGETVRLFEVNGGPNLTSSPH 285
 DB 237 SAGEAVPDVLEAMKTLTPIHIVNGAVGALTGDHALTAAGE--RVLVHSAQNRDTRPH 294
 QY 286 VIGEIFDKVHFEQKGENH--NIOTLLPAGGAATTEKVDVPGDYLVDAIFAAPK 342
 DB 295 LIGCHODV-WATGKFRNPEDLDQETWLTIPGAAGAAYTTPQPGIYAVVNNLIEAFEL 353
 QY 343 GALGILKVEGENHEI 358
 DB 354 GAAGHFYVGDWMDL 369

RESULT 5
 NIR_PSECL STANDARD; PRT; 363 AA.
 ID NIR_PSECL
 AC 006006;

DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
 GN NIK.
 OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 ON NCBI_TaxID=333;
 RX SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.
 RC STRAIN=ATCC 13985;
 RA MEDLINE=93356602; PubMed=8352648;
 RA Glockner A.B., Juegster A., Zumt W.G.;
 RT Copper-containing nitrite reductase from Pseudomonas aureofaciens is
 RT functional in a mutationally cytochrome cdi-free background (Nirs-)
 RT of Pseudomonas stutzeri.";
 RL Arch. Microbiol. 160:18-26(1993).
 CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =
 CC nitrite + ferrocyclochrome c.
 CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE
 CC II COPPER. FAD. TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER,
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
 CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
 CC PSEUDODZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
 CC VITRO.
 CC -1- PATHWAY: Nitrate assimilation (denitrification).
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- DOMAIN: THE TYPE 1 COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
 CC ELECTRON TRANSFER FROM PSEUDODZURIN TO THE TYPE II COPPER SITE
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
 CC REDUCTION OF NITRITE.
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; Z21945; CAA7939.1; .
 CC DR PIR; S32112; S32112.
 CC DR HSSP; P38501; 1A58.
 CC DR InterPro; IPR001117; Cu-oxidase.
 CC DR InterPro; IPR001287; CUNO2_reductase.
 CC DR Pfam; PF00394; Cu-oxidase; 2.
 CC DR PRINTS; PR00695; CUNO2RDTASE.
 CC DR Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
 CC KM Nitrate assimilation; Repeat; Periplasmic; Signal.
 CC FT SIGNAL 1 24
 CC FT CHAIN 25 363
 CC FT DOMAIN 25 193
 CC FT METAL 194 363
 CC FT METAL 113 113
 CC FT METAL 118 118
 CC FT METAL 153 153
 CC FT METAL 154 154
 CC FT METAL 163 163
 CC FT METAL 168 168
 CC FT METAL 324 324
 CC SEQUENCE 363 AA; 39248 MW; 50DDB60CCAD3CB00 CRC64;
 CC
 CC Query Match 17.2%; Score 451; DB 1; Length 363;
 CC Best Local Similarity 39.1%; Pred. No. 1.7e-24;
 CC Matches 124; Conservative 41; Mismatches 116; Indels 36; Gaps 14;
 CC
 CC QY 72 VPPPDROH-----PAKV-VKMETVEKMLAD-GYEYQWRTGGGVGGMIRVRGCD 123
 CC DB 38 VAPPLVHPBOVSGPAPKVVQFRMSIEEKVVIDDQGTLLQMTENGSGPFTLVVHBCD 97

QY 124 TIEVFSNHPDSKMHNDVFNHAATGPGCAEASFTAPCHTSTFSFKALQPLGVYHCAVA 183
 DB 98 YIELTVNPATNSMHNVDVFNHAATGAGALTVVPCQEVLRKADRSSTFFVHC--A 155
 QY 184 PYGM---HIANGWGLLIVERKEGL--PK-----VDKEYVMQGFY-TKRGYEGQLOP 232
 DB 156 PGGMVPMVHVSQMGALMVLPRDGRDPQKLLHYDRVYTTIGESDLTYPKDQGHYKQYP 215
 QY 233 FDMKAIRE-----DAEYVFNFSVGALTGENALAKAYGETVRLFPVNGCPNLTSGF 284
 DB 216 -DLASYDPTRAVMTTLTPSHVFNFRGVALTGANALSKGESV-LPI-HQANDRSIP 272
 QY 285 HVIGEIPDKVHEEGKGEN---HNIQTLLIPAGAAITFEKVDVGVYLVYDHAIFRAN 341
 DB 273 HIGGGHGMV-WTTSKFPANPPQRMNETWFIQGSVALALYFKQGTYYVSHNLIEANE 331
 QY 342 KGALGILVVEGENHEI 358
 DB 332 LGALAQIRVEQWDDDL 348
 CC
 CC RESULT 6
 CC NIK RHIME STANDARD; PRT; 376 AA.
 CC ID NIK RHIME
 CC AC Q92229;
 CC DT 15-SEP-2003 (Rel. 42, Created)
 CC DT 15-SEP-2003 (Rel. 42, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
 CC GN NIK OR R40681 OR SMA1250.
 CC OS Rhizobium meliloti (Sinorhizobium meliloti).
 CC OG Plasmid pSyma (megaplasmid 1).
 CC OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 CC ON NCBI_TaxID=382;
 CC RX SEQUENCE FROM N.A.
 CC RC STRAIN=1021;
 CC RA MEDLINE=21396509; PubMed=11481432;
 CC RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 CC Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 CC Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
 CC Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 CC Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 CC RT "Nucleotide sequence and predicted functions of the entire
 CC Sinorhizobium meliloti pSyma megaplasmid";
 CC Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 CC RL -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =
 CC nitrite + ferrocyclochrome c.
 CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
 CC II COPPER. FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
 CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
 CC PSEUDODZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
 CC VITRO (By similarity).
 CC -1- PATHWAY: Nitrate assimilation (denitrification).
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
 CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
 CC ELECTRON TRANSFER FROM PSEUDODZURIN TO THE TYPE II COPPER SITE
 CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
 CC REDUCTION OF NITRITE.
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE007256; AAK65339.1; .


```

QY 221 ---TKG---KXGEGIQPFMEKAIIE-DAEYVFNQSGALITGENALNAKVEYVRLPV 273
DB 223 PRDEKGEFKKXDSPCGAYEDVAVMRTLPTHTIVFNQAVGALITGENALFAAVE--RVLI 280
QY 274 GNGCPRLTSSFFHYIGISIPDKVHEGKGE-----NNITOTLIPAGAAITEF 321
DB 281 VHSQARBDTRPHIT-----GGHGEVMTGKFNVPVPRDQDTWTFPGPTGAAVY 330
QY 322 KVDVDPDYLVVDHAIIFRAFNKAGILKYEGEENHEIYSHKOTDAYVLEPGA 373
DB 331 TEBQPIYAVVNMNLLEAFELGAAAFKVTGDMNDLMT-----TVRSFSGS 377

RESULT 8
ID NIR_ALCPA STANDARD; PRT; 376 AA.
AC P38501;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR)
GN NIKR OR NIR.
OS Alcaligenes faecalis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_Taxid=511;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-48 AND 192-240.
RC STRAIN=6;
RX MEDLINE=93294530; PubMed=8515232;
RA Nishiyama M., Suzuki J., Kukimoto M., Ohnuki T., Horinouchi S.,
RT Beppu T.;
RT "Cloning and characterization of a nitrite reductase gene from
RT Alcaligenes faecalis and its expression in Escherichia coli.";
RL J. Gen. Microbiol. 139:725-733(1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND MUTAGENESIS.
RC STRAIN=6;
RX MEDLINE=94427056; PubMed=8172899;
RA Kukimoto M., Nishiyama M., Murphy M.E.P., Turley S., Adman E.T.,
RT Horinouchi S., Beppu T.;
RT "X-ray structure and site-directed mutagenesis of a nitrite reductase
RT from Alcaligenes faecalis S-6: roles of two copper atoms in nitrite
RT reduction.";
RL Biochemistry 33:5246-5252(1994).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=96032602; PubMed=7547950;
RA Murphy M.E., Turley S., Kukimoto M., Nishiyama M., Horinouchi S.,
RT Sasaki H., Tanokura M., Adman E.T.;
RT "Structure of Alcaligenes faecalis nitrite reductase and a copper
RT site mutant, M150E, that contains zinc.";
RL Biochemistry 34:12107-12117(1995).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=98019216; PubMed=9353305;
RA Murphy M.E., Turley S., Adman E.T.;
RT "Structure of nitrite bound to copper-containing nitrite reductase
RT from Alcaligenes faecalis. Mechanistic implications.";
RL J. Biol. Chem. 272:28455-28460(1997).
CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferriochrome c =
CC nitrite + ferrocyanochrome c.
CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
CC II COPPER. FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
CC PSEUDOAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
CC VITRO (BY SIMILARITY).
CC -1- PATHWAY: Nitrate assimilation (denitrification).
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: UNDER ANAEROBIC GROWTH CONDITIONS AND BY NITRITE.

```

```

CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
CC ELECTRON TRANSFER FROM PSEUDOAZURIN TO THE TYPE II COPPER SITE
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
CC REDUCTION OF NITRITE.
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D13155; BAA02440.1; -.
CC PIR: I39582; I39582.
CC PDB: 2AFN; 01-AUG-96.
CC PDB: 1NTD; 08-NOV-96.
CC PDB: 1AQ8; 25-FEB-98.
CC PDB: 1AS6; 25-FEB-98.
CC PDB: 1AS7; 25-FEB-98.
CC PDB: 1AS8; 25-FEB-98.
CC PDB: 1ERT5; 24-AUG-00.
CC PDB: 1ERT7; 24-AUG-00.
CC PDB: 1ERT8; 24-AUG-00.
CC DR InterPro: IPR001117; Cu-oxidase.
CC DR InterPro: IPR001287; CuNO2_reductase.
CC DR InterPro: IPR006311; Tac.
CC DR Pfam: PF00394; Cu-oxidase; 2.
CC DR PRINTS: PR00695; CUNO2RDTASE.
CC DR TIGRFAMs: TIGR01409; TAT_signal_seq; 1.
CC KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
CC KW Nitrate assimilation; Repeat; Periplasmic; Signal; 3d-structure;
CC KW Pyroglutamate carboxylic acid.
CC FT SIGNAL 1 33
CC FT CHAIN 34 376
CC FT DOMAIN 34 211
CC FT MOD_RES 212 376
CC FT METIL 131 131
CC FT METAL 136 136
CC FT METAL 171 171
CC FT METAL 172 172
CC FT METAL 181 181
CC FT METAL 186 186
CC FT METAL 342 342
CC FT METAL 171 171
CC FT MUTAGEN 171 186
CC FT MUTAGEN 186
CC -----
CC COPPER-CONTAINING NITRITE REDUCTASE.
CC PLASTOCYANIN-LIKE 1.
CC PYROGLUTAMATE CARBOXYLIC ACID.
CC COPPER (TYPE 1).
CC COPPER (TYPE 2).
CC COPPER (TYPE 2).
CC COPPER (TYPE 1).
CC COPPER (TYPE 1).
CC COPPER (TYPE 1).
CC COPPER (TYPE 1).
CC COPPER (TYPE 2).
CC H->K: LOOSES NITRITE-REDUCING ACTIVITY.
CC M->E: CONTAINS ONLY A TYPE II COPPER ATOM
CC AND FAILS TO CATALYZE THE REDUCTION OF
CC NITRITE.
CC -----
CC HELIX 42 47
CC STRAND 50 52
CC TURN 58 59
CC STRAND 67 67
CC TURN 74 74
CC STRAND 90 91
CC TURN 94 100
CC STRAND 101 102
CC TURN 108 112
CC STRAND 113 113
CC TURN 116 123
CC STRAND 125 126
CC TURN 131 131
CC STRAND 134 135
CC TURN 136 137
CC TURN 141 147
CC HELIX 141 147
CC STRAND 150 150
CC TURN 152 153
CC STRAND 154 161
CC TURN 166 171
CC STRAND 171 171
CC TURN 175 176
CC HELIX 178 182
CC TURN 183 185

```

FT STRAND 187 193
 FT TURN 185 195
 FT STRAND 186 196
 FT STRAND 198 199
 FT TURN 201 202
 FT STRAND 205 206
 FT STRAND 210 219
 FT STRAND 223 222
 FT TURN 225 226
 FT STRAND 229 229
 FT HELIX 235 246
 FT TURN 247 249
 FT STRAND 253 256
 FT TURN 257 258
 FT STRAND 259 259
 FT TURN 260 263
 FT HELIX 265 267
 FT STRAND 269 272
 FT TURN 273 274
 FT STRAND 276 283
 FT STRAND 288 288
 FT STRAND 290 293
 FT TURN 294 294
 FT STRAND 297 301
 FT TURN 302 303
 FT STRAND 305 305
 FT TURN 306 307
 FT STRAND 311 314
 FT STRAND 319 319
 FT TURN 321 322
 FT STRAND 323 330
 FT STRAND 335 341
 FT HELIX 344 348
 FT TURN 349 349
 FT STRAND 353 359
 FT TURN 364 366
 FT STRAND 367 374
 SQ SEQUENCE 376 AA; 40332 MW; 0AD918988301BP29 CRC64;

Query Match 16.3%; Score 427.5; DB 1; Length 376;
 Best Local Similarity 33.7%; Pred. No. 8.1e-23;
 Matches 129; Conservative 51; Mismatches 158; Indels 45; Gaps 14;
 QY 15 LSLMLSGCSNQADKAAQPKSSIVDA---AAKTAANDNAASQEHGCELPVIDAIYTHAP 70
 DB 13 LAGAAALAG-----ALAPVLAITTSAMGOGAVRKATAETALALPRO-KVELVDPPEVYHAAH 64
 QY 71 EVRPYDRDHPAKVVMQETVEKVMRLAD-GVEYQFTFGGOYFGQMI RVRREGDTIEVOF 129
 DB 65 S---QVAEGGPKVVEFTMVEIEKKIYVDDAGTEVHAMAFNGYVPGPLMVVHODDYLELT 121
 QY 130 SNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTFSFALOPGLVYVHCAVAVGM-- 187
 DB 122 INETNTLMNIDPHATGALGGGGLTEINPGEKTIIRFATKRGVGVHC--APPGMPV 179
 QY 188 -HIANGVGLILVEPKGCLP-----KVDKEYVMGSDPYT---KGKGE---QOLOP 232
 DB 180 MHWVSGNNGAIMVLPREGLDHGKALKTYKIYVGGODYVVPDENGKTKKYEAPDAY 239
 QY 233 FDMKAIKE-DAEVVFNFGSVALTGENALKAKGVETVRFVNGNGPNTLTSFHVIGEIF 291
 DB 240 EDVVKVMTLTPPHVFNNGAVALTGKAMTAAGV--KLIIVHSQANRQTRPHLIGCHG 297
 QY 292 DKVHEGSGKGNHNI--QTLIPAGGAITEFKVDVPGDYVLVDHAI FRAFNKGALGILK 349
 DB 298 DYWMATKFTMPRPDQETWTFIPGGAAGAAFYTFQCGIYAVVHNLI EAFELGAAAHFK 357
 QY 350 VEGEENHEIYSHKQDAVYLPEG 372
 DB 358 VTGEMNDLMT-----SVLAPSG 375

RESULT 9

NIR_ALCXX
 ID NIR_ALCXX STANDARD; PRT; 330 AA.
 AC P8145;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Copper-containing nitrite reductase (EC 1.7.2.1) (Cu-NIR).
 GN NIRK.
 OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Achromobacter.
 OX NCBI TaxID=515;
 RN [1]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RC STRAIN=NCIMB 11015;
 RA Dodd F.E., Haenlein S.S., Abraham Z.H.L., Eady R.R., Smith B.E.;
 RT "Structures of a blue-copper nitrite reductase and its substrate-bound complex."
 RL Acta Crystallogr. D 53:406-418(1997).
 RN [2]
 RP CHARACTERIZATION.
 RC STRAIN=NCIMB 11015;
 RX MEDLINE=94059002; PubMed=8240262;
 RA Abraham Z.H.L., Lowe D.J., Smith B.E.;
 RT "Purification and characterization of the dissimilatory nitrite reductase from Alcaligenes xylosoxydans subsp. xylosoxydans (N.C.I.M.B. 11015): evidence for the presence of both type 1 and type 2 copper centres."
 RT Biochem. J. 295:587-593(1993).
 RL -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c = nitrite + ferrocyclochrome c.
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE II COPPER. PAD, TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER, WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
 CC PSEUDODAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR- IN VITRO.
 CC -1- PATHWAY: Nitrate assimilation (denitrification).
 CC -1- SUBUNIT: Homotrimer.
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
 CC -1- DOMAIN: THE TYPE 1 COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR ELECTRON TRANSFER FROM PSEUDODAZURIN TO THE TYPE II COPPER SITE OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE REDUCTION OF NITRITE.
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
 CC PDB; INDS; 07-JUL-97.
 DR PDB; INDS; 07-JUL-97.
 DR InterPro; IPR001117; Cu-Oxidase.
 DR InterPro; IPR001287; CuNO2 reductase.
 DR Pfam; PR00394; Cu-oxidase; 2.
 DR PRINTS; PR00695; CUNO2RDRASR.
 KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; PAD; Nitrite assimilation; Repeat; Periplasmic; 3D-structure.
 FT DOMAIN 1 165
 FT METAL 166 330 PLASTOCYANIN-LIKE 2.
 FT METAL 85 85 COPPER (TYPE 1).
 FT METAL 90 90 COPPER (TYPE 2).
 FT METAL 125 125 COPPER (TYPE 2).
 FT METAL 126 126 COPPER (TYPE 1).
 FT METAL 135 135 COPPER (TYPE 1).
 FT METAL 140 140 COPPER (TYPE 1).
 FT METAL 296 296 COPPER (TYPE 2).
 SQ SEQUENCE 330 AA; 34413 MW; E2C83CA2CEBFCE8 CRC64;

Query Match 16.1%; Score 424; DB 1; Length 330;
 Best Local Similarity 35.5%; Pred. No. 1.2e-22;
 Matches 116; Conservative 36; Mismatches 119; Indels 56; Gaps 11;
 QY 72 VRRPVDRD-----PARVVMQETVEKVMRL-ARGVEYQFTFGGOYFGQMI RVRREGD 123
 DB 10 VAPPLVHPHSQVAAAGAPKVVQPFMSIEKKVMADDDGTTAAQAMTENGSPGRLTVVHGED 69
 QY 124 TIEVQFSNHPDSKMPHNVDFHAATGPGGAASFTAPGHTSTFSFALOPGLVYVHCAVA 183

```

Db      70 YIELTVNATSMFPHVDFHAATGALGGAGLQVVGEGDAVREFADRSGETFVHC--A 127
Qy      184 PVGM--HIANGMYGLIIVEPKKGLPKV-----DKERYVMQGPY-----TKKXGE-- 227
Db      128 PAGWVPMHVSGMNGMLVLPDGLRDAGALAYRYVTIGESDLYVPKADGNSDVP 187
Qy      228 -----QGIQPFDMKAIREDAEYVNGSVGALTGEMAKKAGVETVRLFVG 274
Db      188 ALASAVADTVAVMRTLP-----SHAVFNGAVAGALTGANALTAAGVESVLI-I-- 234
Qy      275 NGSPVLTSFFHYIGIFDKVHFEKGGENH---NIOTLILPAGGAITEFKVDVPDYVL 331
Db      235 HSOARDSKPHLIGHGDMV-WTTGKFAHPOLNMTWPIPGSAAALYTRKQPGTVAY 293
Qy      332 VDAIFRAFNKAGALGILKVEGSENHET 358
Db      294 LSHNLIEMELGAAQAQASVEGQWDDDL 320

RESULT 10
PCOA_ECOLI
ID      PCOA_ECOLI      STANDARD;      PRT;      605 AA.
AC      Q4752;
DT      16-OCT-2001 (rel. 40, Created)
DT      16-OCT-2001 (rel. 40, Last sequence update)
DT      16-OCT-2001 (rel. 40, Last annotation update)
DE      Copper resistance protein A precursor.
GN      PCOA.
OS      Escherichia coli.
OC      Bacteriia; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / ED8739;
RX      MEDLINE=96130847; PubMed=8594334;
RA      Brown N.L., Barrett S.R., Camakaris J., Lee B.T.O., Rouch D.A.;
RT      "Molecular genetics and transport analysis of the copper-resistance
RL      Mol. Microbiol. 17:1153-1166(1995).
CC      - FUNCTION: REQUIRED FOR THE COPPER-INDUCIBLE EXPRESSION OF COPPER
CC      - SUBCELLULAR LOCATION: Periplasmic (Potential).
CC      - SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES. COPA
CC      SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, X83541; CAA58525.1; -.
DR      PIR, S70159; S52253.
DR      HSSP, P36649; 1K77.
DR      InterPro, IPR006376; COPA.
DR      InterPro, IPR001117; Cu-oxidase.
DR      InterPro, IPR002355; MultiCu_oxidase2.
DR      InterPro, IPR006311; Tat.
DR      Pfam, PF00394; Cu-oxidase; 2.
DR      TIGRfams, TIGR01480; copper_res_A; 1.
DR      TIGRfams, TIGR01409; tat_signal_seq; 1.
DR      PROSITE, PS00079; MULTICOPPER_OXIDASE1; 1.
DR      PROSITE, PS00080; MULTICOPPER_OXIDASE2; 1.
KW      Oxidoreductase; Copper; Repeat; Signal; Plasmid; Metal-binding;
KW      Periplasmic.
FT      SIGNAL      1      40
FT      CHAIN      41      605
FT      DOMAIN      382      429      3 X 8 AA TANDEM REPEATS OF D-H-X-X-M-X-

```

```

FT      REPEAT      382      389      G-M.
FT      REPEAT      414      421      1.
FT      REPEAT      422      429      2.
FT      METAL      100      100      3.
FT      METAL      102      102      COPPER (TYPE 2) (BY SIMILARITY).
FT      METAL      142      142      COPPER (TYPE 3) (BY SIMILARITY).
FT      METAL      144      144      COPPER (TYPE 3) (BY SIMILARITY).
FT      METAL      144      144      COPPER (TYPE 3) (BY SIMILARITY).
FT      METAL      538      538      COPPER (TYPE 1) (POTENTIAL).
FT      METAL      541      541      COPPER (TYPE 2) (POTENTIAL).
FT      METAL      543      543      COPPER (TYPE 3) (POTENTIAL).
FT      METAL      586      586      COPPER (TYPE 3) (POTENTIAL).
FT      METAL      587      587      COPPER (TYPE 1) (POTENTIAL).
FT      METAL      588      588      COPPER (TYPE 3) (POTENTIAL).
FT      METAL      592      592      COPPER (TYPE 1) (POTENTIAL).
FT      METAL      597      597      COPPER (TYPE 1) (POTENTIAL).
SQ      SEQUENCE      605 AA; 67307 MW; 8EECA182D56B27D8 CRC64;

Query Match      7.7%; Score 203.5; DB 1; Length 605;
Best Local Similarity 22.4%; Pred. No. 8e-07;
Matches 89; Conservative 88; Mismatches 167; Indels 83; Gaps 14;

Qy      100 GVEYQVTFPGGQVPGQMTVRREGDTIEVQFSNHPDSK-----MHNVDFFHAATGCG 150
Db      60 GSERQAKTINGLPGPVLRMRKGDITTLKVNRLNEQTSIHWHGIIIPAND----- 111
Qy      151 GGAASFTAA--PGHTSPFSKALOPGLVYVCAVAPVGMHIANGYGLIIVEPKKGLP-X 207
Db      112 GVPGLSFGIGIEPDDTYVTFKVKQNGTYWH--SHSGLOEGGVGALIIIDAREPEPA 168
Qy      208 VDKEYVYVQGDFFYTKGKYGEQGLQPFDMKAIREDAEYVFN---GSGVALTGENALKA 263
Db      159 YDRHIVVLSWTD-----NPHSLKKLKKQSDYVNFNKPVGSFPRVNRGLSA 220
Qy      264 KVGETVRLFVNGGNGNLTSFHYIGEIDPKYHFBGCKENHNIQTTLPA-----CG 315
Db      221 TIAD--RQWMAEMKNPFDLADVSG--YTYTYLWNGQAPLKNMTGLFPRGKIRLRFTNG 276
Qy      316 AATFEFKVDVGDYVLYVHAIFRAFNKAGALGILKVEGSENHETYSKOTDAVYL----- 369
Db      277 SMATYFDIRIRGLKKTVAALAQGYVNPVTVEFRFAVETDYVVEPGGEATITTAQSMQ 336
Qy      370 -----PEGAPQAIIDQEAQKTPAPANLQEQIKAGATYD-----SNC 406
Db      337 RTGVARGLATREGISAIV-----PLDRPLILTMEDMGMGCHDMAGMDHSQMGMDNS 392
Qy      407 AACQPDGKGVN--AFPLANSDYLNADHARAASI 440
Db      393 GEMMSMDGADLPDGSSTSPMDHSSMAGMDHSRMAGM 429

RESULT 11
ASO_CUCPM
ID      ASO_CUCPM      STANDARD;      PRT;      552 AA.
AC      P37064;
DT      01-JUN-1994 (rel. 29, Created)
DT      01-JUN-1994 (rel. 29, Last sequence update)
DT      28-FEB-2003 (rel. 41, Last annotation update)
DE      L-ascorbate oxidase (EC 1.10.3.3) (Ascorbate) (ASO).
OS      Cucurbita pepo var. melopepo (Zucchini).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eutrosids I; Cucurbitales; Cucurbitaceae; Cucurbita.
OX      NCBI_TaxID=3665;
RN      [1]
RP      X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX      MEDLINE=92194315; PubMed=1548698;
RA      Messerschmidt A., Ladenstein R., Huber R., Bolognesi M.,
RA      Avigliano L., Petruzzelli R., Rossi A., Finazzi-Agro A.;
RT      "Refined crystal structure of ascorbate oxidase at 1.9-A resolution.";
RL      J. Mol. Biol. 224:179-205 (1992).
RN      [2]
RP      X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

```


RX MEDLINE=89236417; PubMed=2716059;
RA Meserich A., Rosen A., Ladenschneider R., Huber R., Bolognesi M.,
RT Gatti G., Marchesini A., Petrucci R., Finazzi-Aro A.,
RT "X-ray crystal structure of the blue oxidase ascorbate from
RT zucchini. Analysis of the polypeptide fold and a model of the copper
RT sites and ligands.";
RL J. Mol. Biol. 206:513-529 (1989).
CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC
CC ACID.
CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2
CC H(2)O.
CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: contains 3 plastocyanin-like domains.
DR PIR: A51027; A51027.
DR PDB: 1A0Z; 31-OCT-93.
DR PDB: 1ASQ; 31-JAN-94.
DR PDB: 1ASQ; 31-JAN-94.
DR GLYCOSITE: P37064; -.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR002355; MultiCu-oxidase2.
DR Pfam: PF00394; Cu-oxidase; 3-
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
KW Oxidoreductase; Copper; Glycoprotein; 3D-structure; Repeat.
FT DOMAIN 1 122 PLASTOCYANIN-LIKE 1.
FT DOMAIN 134 300 PLASTOCYANIN-LIKE 2.
FT DOMAIN 344 523 PLASTOCYANIN-LIKE 3.
FT DISULFID 19 201
FT DISULFID 81 538
FT DISULFID 180 193
FT CARBOHYD 92
FT METAL 60 60 N-LINKED (GLCNAC. . .).
FT METAL 62 62 /FTID=CAR.000149.
FT METAL 104 104 COPPER (TYPE 2).
FT METAL 106 104 COPPER (TYPE 3).
FT METAL 445 445 COPPER (TYPE 3).
FT METAL 448 445 COPPER (TYPE 1).
FT METAL 448 448 COPPER (TYPE 2).
FT METAL 450 450 COPPER (TYPE 3).
FT METAL 506 506 COPPER (TYPE 3).
FT METAL 507 507 COPPER (TYPE 1).
FT METAL 508 508 COPPER (TYPE 1).
FT METAL 512 512 COPPER (TYPE 3).
FT METAL 517 517 COPPER (TYPE 1).
FT METAL 517 517 COPPER (TYPE 1).
FT STRAND 3 15
FT TURN 17 18
FT STRAND 22 27
FT TURN 28 29
FT STRAND 30 30
FT STRAND 30 30
FT STRAND 36 39
FT TURN 40 41
FT STRAND 43 50
FT STRAND 57 57
FT STRAND 57 57
FT STRAND 59 62
FT TURN 63 63
FT TURN 63 63
FT TURN 67 68
FT HELIX 70 72
FT STRAND 76 76
FT STRAND 77 79
FT TURN 80 80
FT STRAND 83 83
FT STRAND 83 83
FT TURN 85 86
FT STRAND 94 94
FT STRAND 99 105
FT TURN 108 110
FT TURN 110 110
FT HELIX 111 113
FT TURN 114 114
FT STRAND 116 122
FT TURN 125 126

FT STRAND 135 144
FT HELIX 149 155
FT TURN 156 157
FT STRAND 169 172
FT TURN 173 174
FT STRAND 175 175
FT STRAND 181 181
FT STRAND 182 183
FT TURN 184 186
FT HELIX 188 189
FT TURN 192 192
FT STRAND 200 201
FT TURN 206 208
FT STRAND 210 211
FT TURN 213 220
FT STRAND 226 231
FT STRAND 232 233
FT TURN 236 241
FT STRAND 242 243
FT STRAND 244 251
FT STRAND 254 256
FT TURN 258 259
FT STRAND 261 267
FT TURN 272 273
FT STRAND 276 283
FT STRAND 292 298
FT TURN 299 300
FT TURN 303 304
FT TURN 317 318
FT HELIX 320 327
FT TURN 328 329
FT STRAND 332 332
FT STRAND 334 335
FT STRAND 345 356
FT TURN 357 358
FT STRAND 359 364
FT TURN 367 369
FT STRAND 376 381
FT HELIX 382 383
FT TURN 385 386
FT TURN 399 400
FT TURN 403 404
FT TURN 410 411
FT STRAND 413 415
FT STRAND 419 421
FT TURN 423 424
FT STRAND 426 433
FT TURN 439 440
FT STRAND 445 449
FT TURN 450 451
FT STRAND 454 461
FT HELIX 466 471
FT STRAND 479 480
FT STRAND 482 485
FT TURN 487 488
FT STRAND 489 496
FT STRAND 501 507
FT STRAND 510 514
FT HELIX 515 516
FT TURN 518 523
FT STRAND 525 527
FT HELIX 533 536
FT HELIX 539 545
SQ SEQUENCE 552 AA; 61704 MW; 24660B0F47AB54B4 CRC64;

Query Match

Best Local Similarity 6.5%; Score 170; DB 1; Length 552;

Matches 81; Conservative 47; Mismatches 129; Indels 80; Gaps 18;

QY 101 VEQVFT-----FGGVPGQIRVREGDTIVQFSN--HPDSKMPHNVDFPAAG 148
Db 10 VEFYFWAPNCNENIVMGINGQFPFGPTIRANAGDSVVELTNKLHTEGVVH--WHGILO 66

```

QY 149 PG-----GGAEASFTL--PGHSTSPFKALQPGLYVYHCAVAPVGMHIANMVGILVLP 201
DB 67 RGTPMADGTASISQCAINNGEFTFFNFVNDNPGTFEYH---GHLMQORAGLGLSLVDP 123
QY 202 KEGLEPK---VDKEYVWMOGDFYTKGKYGGQ-GLOPFDMKATP--EDAEVYVNG----- 250
DB 124 POGKEPFRHYDEINLLSDMWHQSIHKEVGLS-----SKPIRWIGEPOTILLNGRGQD 179
QY 251 -SVGALTGEN--ALKAKVGETVRLFYNGGPNLTSSFHVIGEIFDKVHEGKGGENHNQ 307
DB 180 CSIAKYDSNLEBCKLKGSESCAPYIFHVSFKTYRIRI-----AS 220
QY 308 TLLIPAGAAITEFK---VDVPGDYV---LVDAIFRAFNKALGILKEGGENHEI-- 358
DB 221 TLLALALNFAIGNHQLLVVEADGNVQPFYTSIDIDYSGESYSVLITTDQNPSENWVSV 280
QY 359 ---SHKQTD-----AVYLPBGAPOAIDQEAAPKTPA 387
DB 281 GTRARHPNTPPGTLTLNLYLPNSVSK-LPTSPPEPTPA 316

RESULT 12
ASO_CUCMA STANDARD; PRT; 579 AA.
ID P24792; Q39539;
AC 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-ascorbate oxidase precursor (EC 1.10.3.3) (Ascorbase) (ASO).
GN AAO.
OS Cucurbita maxima (Pumpkin) (Winter squash).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid I; Cucurbitales; Cucurbitaceae; Cucurbita.
OK NCBI_Taxid=3661;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-48.
RC STRAIN=cv. Ebisu Nankin;
RX MEDLINE=90361033; PubMed=2143984;
RT Esaka M., Harori T., Fujisawa K., Sakajo S., Asahi T.;
RT "Molecular cloning and nucleotide sequence of full-length cDNA for
RT ascorbate oxidase from cultured pumpkin cells.";
RT Eur. J. Biochem. 191:537-541(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97354114; PubMed=9210335;
RT Kisu Y., Harada Y., Goto M., Esaka M.;
RT "Cloning of the pumpkin ascorbate oxidase gene and analysis of a cis-
RT acting region involved in induction by auxin.";
RT Plant Cell Physiol. 38:631-637(1997).
CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC
CC ACID.
CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2
CC H(2)O.
CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
CC 2 OR NORML, AND TYPE 3 OR COUPLED BINUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X55779; CAA39300.1; -
CC DR EMBL; D55677; BAA09528.1; -
CC DR HSSP; P37064; IAOZ.
CC InterPro; IPR001117; Cu-oxidase.

```

```

DR InterPro; IPR002355; MultiCu_oxidase2.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
KW Oxidoreductase; Copper; Metal-binding; Glycoprotein; Signal; Repeat.
FT SIGNAL 1 30
FT CHAIN 31 579
FT DOMAIN 33 152 L-ASCORBATE OXIDASE.
FT DOMAIN 164 330 PLASTOCYANIN-LIKE 1.
FT DOMAIN 374 553 PLASTOCYANIN-LIKE 2.
FT DISULFID 49 231 PLASTOCYANIN-LIKE 3.
FT DISULFID 111 568 BY SIMILARITY.
FT DISULFID 210 223 BY SIMILARITY.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT METAL 90 90 COPPER (TYPE 2) (PROBABLE).
FT METAL 92 92 COPPER (TYPE 3) (PROBABLE).
FT METAL 134 134 COPPER (TYPE 3) (PROBABLE).
FT METAL 136 136 COPPER (TYPE 3) (PROBABLE).
FT METAL 475 475 COPPER (TYPE 1) (PROBABLE).
FT METAL 478 478 COPPER (TYPE 2) (PROBABLE).
FT METAL 480 480 COPPER (TYPE 3) (PROBABLE).
FT METAL 536 536 COPPER (TYPE 3) (PROBABLE).
FT METAL 537 537 COPPER (TYPE 1) (PROBABLE).
FT METAL 538 538 COPPER (TYPE 3) (PROBABLE).
FT METAL 542 542 COPPER (TYPE 1) (PROBABLE).
FT METAL 547 547 COPPER (TYPE 1) (PROBABLE).
FT CONFLICT 175 175 W -> C (IN REF. 1) (POTENTIAL).
SQ SEQUENCE 579 AA; 64667 MW; 8F5AF4CB07B27689 CRC64;

Query Match 6.4%; Score 167; DB 1; Length 579;
Best Local Similarity 24.6%; Pred. No. 0.00027;
Matches 83; Conservative 45; Mismatches 129; Indels 80; Gaps 19;

QY 101 VEYQVWT-----FEGVPGQMIRVEDTIEVQFSN--HPDSKPHNVDFHATG 148
DB 40 VEYMFAPDCNENIWMGINGQFPETIRANADTVVELINKLHTEGVYH---WHGIQ 96
QY 149 PG-----GGAEASFTL--PGHSTSPFKALQPGLYVYHCAVAPVGMHIANMVGILVLP 201
DB 97 RGTPMADGTASISQCAINNGEFTFFNFVNDNPGTFEYH---GHLMQORAGLGLSLVDP 153
QY 202 KEGLEPK---VDKEYVWMOGDFYTKGKYGGQ-GLOPFDMKATP--EDAEVYVNG----- 250
DB 154 POGKEPFRHYDEINLLSDMWHQSIHKEVGLS-----SKPIRWIGEPOTILLNGRGQD 209
QY 251 -SVGALTGEN--ALKAKVGETVRLFYNGGPNLTSSFHVIGEIFDKVHEGKGGENHNQ 307
DB 210 CSIAKYDSNLEBCKLKGSESCAPYIFHVSFKTYRIRI-----AS 250
QY 308 TLLIPAGAAITEFK---FVYDVPDYV---LVDAIFRAFNKALGILKEGGENHEI-- 359
DB 251 TLLALALNFAIGNHPLLVVEADGNVQPFYTSIDIDYSGESYSVLITTDQNPSENWVSV 310
QY 360 ---SHKQTD-----AVYLPBGAPOAIDQEAAPKTPA 387
DB 311 GTRGRHPNTPPGTLTLNLYLPNSVSK-LPTSPPEPTPA 346

RESULT 13
CPAL_PSESM STANDARD; PRT; 609 AA.
ID CPAL_PSESM
AC P12374;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper resistance protein A precursor.
GN COPA.
OS Pseudomonas syringae (pv. tomat).
OC Plasmid pPT23D.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.

```

OX NCBI_TaxId=323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PT23.2;
 RA MEDLINE=88227880; PubMed=3372485;
 RX Mellano M.A., Cooksey D.A.;
 RT "Nucleotide sequence and organization of copper resistance genes from
 RL *Pseudomonas syringae* pv. tomato.";
 RN *J. Bacteriol.* 170:2879-2883(1988).
 [2]
 RP CHARACTERIZATION, AND SEQUENCE OF 33-37.
 RC STRAIN=PT23.2;
 RX MEDLINE=92020961; PubMed=1924351;
 RA Cha J.-S., Cooksey D.A.;
 RT "Copper resistance in *Pseudomonas syringae* mediated by periplasmic and
 RL outer membrane proteins.";
 CC Proc. Natl. Acad. Sci. U.S.A. 88:8915-8919(1991).
 CC -1- FUNCTION: MEDIATES COPPER RESISTANCE BY SEQUESTRATION OF COPPER IN
 CC THE PERIPLASM ALONG WITH THE COPPER-BINDING PROTEIN COPC. MAY HAVE
 CC OXIDASE ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY COPPER.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES. COPA
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1eb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL_M19930; AAA25806.1; -.
 DR FIR; A32018; KSPSCY.
 DR HSSP; P36649; 1KV7.
 DR InterPro: IPR006376; COPA.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; MulticCu-oxidase2.
 DR InterPro: IPR006311; Tat.
 DR Pfam; PF00394; Cu-oxidase; 2.
 DR TIGRFAMS; TIGR01480; copper_res_A; 1.
 DR TIGRFAMS; TIGR01409; Tat signal_seq; 1.
 DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
 DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
 DR Oxidoreductase; Copper; Repeat; Signal; Plasmid; Metal-binding;
 KM Periplasmic.
 KW
 FT SIGNAL 1 32
 FT CHAIN 33 609 COPPER RESISTANCE PROTEIN A.
 FT DOMAIN 367 434 5 X 8 AA TANDEM REPEATS OF D-H-X-X-M-X-
 FT 367 374 1.
 FT REPEAT 375 382 2.
 FT REPEAT 408 415 3.
 FT REPEAT 419 426 4.
 FT REPEAT 427 434 5.
 FT METAL 100 100 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 102 102 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 142 142 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 144 144 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 542 542 COPPER (TYPE 1) (POTENTIAL).
 FT METAL 545 545 COPPER (TYPE 2) (POTENTIAL).
 FT METAL 547 547 COPPER (TYPE 3) (POTENTIAL).
 FT METAL 590 590 COPPER (TYPE 3) (POTENTIAL).
 FT METAL 591 591 COPPER (TYPE 1) (POTENTIAL).
 FT METAL 592 592 COPPER (TYPE 3) (POTENTIAL).
 FT METAL 596 596 COPPER (TYPE 1) (POTENTIAL).
 FT METAL 601 601 COPPER (TYPE 1) (POTENTIAL).
 SO SEQUENCE 609 AA; 67354 MW; 9C086528B8135252 CRC64;

QY 110 GQVPGMIRVREEDTIEVQSNHPDSK-----MPHNVDFAATGPGGAASFTA- 159
 Db 70 GGLPGLRLKWKGDVTLVRRLDAATSIMHGIILPPNM-----GVPGLSFGAI 121
 QY 160 -DGHSTSTSEFKALQGLVYHCAVAPVGMHIANMGLIVPEKGLP-KVDKEYVVMG 217
 Db 122 EPGGVYVQFKVQOAGTWYH---SHSGEQEVGVPLVLEAKPEPEPKYDSEHVMTL 178
 QY 218 DEYTGKVGEOGLQEPDEKAIRDEAEVYFNGSGALTGEMALAKAGETRLFFVNGG 277
 Db 179 DWTDE-----DPSLRLTKKOSDYNEH-----KRYGDFVNDVADKCM 218
 QY 278 P-----NLTSFVHVEIPEPKVFEKSGKGNHNIQTLLPA-----GG 315
 Db 219 AATVADRKMMAMKKNPTDLADVSGATY--TYLLAQAPNMWMTGLFRGKTLRLPIFG 276
 QY 316 AATFERKVDVPGDYVLVDHAIIRAFNKGALGILKVBGEENHEIYHKQTDVAYLPEGAPQ 375
 Db 277 SAMTYFDIRIPGLKMTVVASDQFVNPVEVDLRIAVETPVIPEPTAEAYTV--FAQ 333
 QY 376 AID-----TOEAPKTPAPANLQEQIKKATY-----DSNC 406
 Db 334 SMDRGTGARGTLAVBEGLVAQVPLDPRPLVTMDMGKGMDSMGDSMGDSGADGCM 393
 QY 407 AACHPDQKGVNAPPLANSDYLNADHARAASIVANGISGKI-----TVNGN---QYE 457
 Db 394 QTMSSMGDSMNP-ANDHSMGSTMGMDSMGDSMGDSMGDSMGDSMNSHPASENDPLVDWG 452
 QY 458 SYMPALISDQQA-----NVITY 476
 Db 453 AMSPTAKLNDPGLGRNNGRKVLT 477
 RESULT 14
 ID CYCA_GLUOX STANDARD; PRT; 478 AA.
 AC 047945; 008083;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alcohol dehydrogenase cytochrome c subunit precursor (Cytochrome c-
 DE 553) (C553) (G3-ADH subunit II).
 GN ADHB.
 OS Gluconobacter oxydans (Gluconobacter suboxydans).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Acetobacteraceae; Gluconobacter.
 OC NCBI_TaxId=442;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Takeda Y., Shimizu T.;
 RT "Cloning and sequencing of the gene encoding cytochrome c-553 (CO)
 RL from *Gluconobacter suboxydans*.";
 RL U. Ferment. Bioeng. 72:1-6(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-52.
 RC STRAIN=IFO 12528;
 RX MEDLINE=97208225; PubMed=9055427;
 RA Kondo K., Horiouchi S.;
 RT "Characterization of the genes encoding the three-component membrane-
 bound alcohol dehydrogenase from *Gluconobacter suboxydans* and their
 RT expression in *Acetobacter pasteurianus*.";
 RL Appl. Environ. Microbiol. 63:1131-1138(1997).
 CC -1- FUNCTION: HETEROTRIMER (DEHYDROGENASE, CYTOCHROME AND PROTEIN
 CC ADHS), THAT FORMS THE ALCOHOL DEHYDROGENASE COMPLEX.
 CC -1- SUBUNIT: HETEROTRIMER (DEHYDROGENASE, CYTOCHROME AND PROTEIN
 CC ADHS), THAT FORMS THE ALCOHOL DEHYDROGENASE COMPLEX.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND, FACING THE PERIPLASMIC
 CC SPACE (POTENTIAL).
 CC -1- PTM: BINDS THREE HEME GRUUPS PER MOLECULE (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M58760; AAA4935.1; -
EMBL; D6375; BAA19754.1; -
HSSP; P04164; 1DT1.
InterPro; IPR003088; Cyt_C1.
InterPro; IPR002329; Cyt_C1C.
InterPro; IPR003219; Cyt_C1C.
InterPro; IPR000345; CytC_heme_bind.
Pfam; PF00034; Cytochrome_c_1.
PRINTS; PR00605; CYTCROME1C.
ProDom; PD011584; CytC adh; 1.
PROSITE; PS00190; CYTOCHROME_C_3.
Electron transport; Oxidoreductase; Signal; Heme; Membrane; Periplasmic; Pyrrolidone carboxylic acid.
KW Periplasmic; Pyrrolidone carboxylic acid.
FT SIGNAL 1 36
FT CHAIN 37 478
FT MOD_RES 37 37 ALCOHOL DEHYDROGENASE CYTOCHROME C
FT BINDING 56 56 PYRROLIDONE CARBOXYLIC ACID.
FT BINDING 59 59 HEME 1 (COVALENT) (BY SIMILARITY).
FT METAL 60 60 HEME 1 (COVALENT) (BY SIMILARITY).
FT BINDING 204 204 HEME 2 (COVALENT) (BY SIMILARITY).
FT BINDING 207 207 HEME 2 (COVALENT) (BY SIMILARITY).
FT METAL 208 208 IRON (HEME 2 AXIAL LIGAND) (BY SIMILARITY).
FT BINDING 340 340 HEME 3 (COVALENT) (BY SIMILARITY).
FT BINDING 343 343 HEME 3 (COVALENT) (BY SIMILARITY).
FT METAL 344 344 IRON (HEME 3 AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 478 AA; 51198 MW; 3FBC1F935AD1D866 CRC64;

Query Match 6.2%; Score 163; DB 1; Length 478;
Best Local Similarity 36.0%; Pred. No. 0.00041;
Matches 49; Conservative 18; Mismatches 59; Indels 10; Gaps 6;

QY 369 LPEGAPQADTDDEAPRTAPPAULOBIKAGKATYDSNCAHQPPDKGVNAPPLA-NS 427
DB 306 VEGKNLG---ODDQGTTLALNKGGGNAGAEVYLNCALCHMNDGTGNRWFPLPAGNP 362
QY 428 DYLNADHRAASIVANGSGKITVNGQYEV-MPALA--LSDQOIANYITLINSFGNK 484
DB 363 VITTDPTSLAVAVFG--GILPTNSAPSAVAMPFGKNHLSQEMADVVFARKGMGN 420
QY 485 G-GQSSADDAVAKAKT 499
DB 421 APTGVSASDIQKLRFT 436

RESULT 15
ASO_TOBAC STANDARD; PRT; 578 AA.
ID_ASO_TOBAC
AC Q40588;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-UL-1999 (Rel. 38, Last annotation update)
DE L-ascorbate oxidase precursor (EC 1.10.3.3) (Ascorbase) (ASO).
GN AAO.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96194464; PubMed=8624413;
RA Kato N., Esaka M.;
RT "cDNA cloning and gene expression of ascorbate oxidase in tobacco";
RL Plant. Mol. Biol. 30:833-837(1996).

CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC
CC ACID.
CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2
CC H(2)O.
CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN YOUNG AND GROWING TISSUES.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; D43624; BAA07734.1; -
PIR; S66353; S66353.
HSSP; P37064; 1AO2.
InterPro; IPR001117; Cu-oxidase.
InterPro; IPR002355; Multicopper oxidase2.
Pfam; PF00394; Cu-oxidase; 3.
PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
KW Signal; Oxidoreductase; Copper; Glycoprotein; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 578
FT DOMAIN 30 149 L-ASCORBATE OXIDASE.
FT DOMAIN 161 328 PLASTOCYANIN-LIKE 1.
FT DOMAIN 372 550 PLASTOCYANIN-LIKE 2.
FT DISULFID 108 565 PLASTOCYANIN-LIKE 3.
FT DISULFID 207 221 BY SIMILARITY.
FT METAL 87 87 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 89 89 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 131 131 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 133 133 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 472 472 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 475 475 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 477 477 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 533 533 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 534 534 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 535 535 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 539 539 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 544 544 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 438 438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 578 AA; 64865 MW; C2371F5FEA353E3B CRC64;

Query Match 6.2%; Score 162; DB 1; Length 578;
Best Local Similarity 22.9%; Pred. No. 0.00061;
Matches 104; Conservative 55; Mismatches 162; Indels 134; Gaps 23;

QY 91 VEKVRNLADGYEYQFWTEGGQVPGQMIIVREEDTIEVQFSN--HPSKMPHNVDFHATG 148
DB 37 VEYIHWSPDGEESVVMGJNGQFPGPPIRAKAGDTVAHLTNKLTGEGVYH---WGIRO 93
QY 149 PG-----GGAASFLA--PGHTSTSFKALDQGLVYVNCVAVPVMHIANGVYGLIVRP 201
DB 94 IGTTPADGTALISQCAINPGETFLRFVYDAGTYFYH--GHYQMOSAGIYGLSILVEV 150
QY 202 KEGLPK--VDKEYVWGDFTYTKGYEGOGLOPFDEMEKATREDAVYVFNFGVGLNGE 258
DB 151 GEGEKEKPFHYDGEFLLSLDWMHKGSH-----QEVLD-----SS 185
QY 259 NALKAKVGETVRLVY-GNGGPNLTSSFVIGEIFDKVHFEGGKGNHNI----- 306
DB 186 NPLR-WIEGPOTLLNGRGQYVNCISLAARFSPRLPQCKLRGGEQVAPQILRVPRNKIYRL 244

```

QY 307 ---QTLIPAGCAITEFK---VDVPGDYLVDAIFRAFNKCALGLIKVGESENHEIYS 360
DB 245 RVASTIALSLAIGCHKVNVVADGNVY-----QPS-----VQMDIYS 286
QY 361 HKQTDVYLPEGAPO-----AIDTO-EAPKTPANLQEOIKAKATYDSCNCAHQPDG 414
DB 287 GESYSVLFKTDQPTKNWISIVNRREPTPGGLT-----LNYLNSAS----- 332
QY 415 KGVNPAFPLAN--SDYLNADHRAASIVANGLSGK-----ITYNG----- 453
DB 333 -KFTLPPLPIAPIMNDY---NHSKSPSNKIFALMGSPKPPQNHRIILNTONKIDGYT 388
QY 454 ---NGYESVPAIALSDQOIANVITYTTLNSFGNK 484
DB 389 KMAINNVSLVLPF-----QVYLGSIKYGIAFDTK 418

RESULT 16
C552_THETH STANDARD; PRT; 148 AA.
ID P04164;
AC 20-MAR-1987 (Rel. 04, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c-552 precursor (C552).
GN CYC.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxId=274;
RN [1]
RP SEQUENCE FROM N.A.
RA Keightley J.A., Wacher M.W., Fee J.A.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-148.
RC STRAIN=HB8 / ATCC 27634;
RA MEDLINE=8519131; PubMed=2986625;
RT Titani K., Ericsson L.H., Hon-Nami K., Miyazawa T.;
RT "Amino acid sequence of cytochrome c-552 from Thermus thermophilus
   HB8."
RL Biochem. Biophys. Res. Commun. 128:781-787(1985).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.28 ANGSTROMS).
RA MEDLINE=97428333; PubMed=9281430;
RT Than M.E., Hof P., Huber R., Bourenkov G.P., Bartunik H.D., Buse G.,
RA Soulimane T.;
RT "Thermus thermophilus cytochrome-c552: a new highly thermostable
   cytochrome-c structure obtained by MAD phasing."
RL J. Mol. Biol. 271:629-644(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 20-148.
RC STRAIN=HB8 / ATCC 27634;
RA MEDLINE=21025441; PubMed=1152119;
RT Fee J.A., Chen Y., Todaro T.R., Bren K.L., Patel K.M., Hill M.G.,
RA Gomez-Moran E., Loshr T.M., Ai J., Thony-Meyer L., Williams P.A.,
RA Stura E., Stridhar V., McKee D.E.;
RT "Integrity of Thermus thermophilus cytochrome c552 synthesized by
   Escherichia coli cells expressing the host-specific cytochrome c
   maturation genes, ccmABCDGHI: biochemical, spectral, and structural
   characterization of the recombinant protein."
RL Protein Sci. 9:2074-2084(2000).
RN [5]
RP FUNCTION: THIS MONOHEME BASIC PROTEIN APPEARS TO FUNCTION AS AN
   ELECTRON DONOR TO CYTOCHROME OXIDASE IN T.THERMOPHILUS. IT MAY
   BE A C2-TYPE CYTOCHROME BUT IS DISTANTLY RELATED TO OTHER
   CYTOCHROMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: M93437; AAB8580.1; -.
DR PDB: 1C52; 24-JUN-98.
DR PDB: 1DT1; 28-MAR-01.
DR PDB: 1FOC; 09-OCT-02.
DR InterPro: IPR003088; Cyt C1.
DR InterPro: IPR003045; CytC_heme_bind.
DR Pfam: PF00034; cytochrome_c_1.
DR PROSITE: PS00190; CYTOCHROME_C_1.
KW Electron transport; Heme; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 148 CYTOCHROME C-552.
FT BINDING 28 28 HEME (COVALENT).
FT BINDING 31 31 HEME (COVALENT).
FT METAL 32 32 IRON (HEME AXIAL LIGAND).
FT METAL 80 80 IRON (HEME AXIAL LIGAND) (OR M-86).
FT HELIX 21 24
FT HELIX 25 25
FT TURN 26 32
FT TURN 34 35
FT STRAND 39 39
FT TURN 40 42
FT STRAND 43 43
FT TURN 47 48
FT HELIX 49 54
FT TURN 55 55
FT TURN 57 58
FT HELIX 59 69
FT STRAND 71 77
FT TURN 78 79
FT STRAND 80 86
FT TURN 90 91
FT HELIX 94 106
FT TURN 107 108
FT TURN 110 110
FT HELIX 111 113
FT TURN 115 116
FT HELIX 122 128
FT TURN 129 130
FT HELIX 135 143
FT TURN 144 145
SQ SEQUENCE 148 AA; 15974 MW; 6AE7E806BDB5EALD CRC64;

Query Match 6.1%; Score 161; DB 1; Length 148;
Best Local Similarity 37.3%; Pred. No. 0.00012;
Matches 44; Conservative 23; Mismatches 39; Indels 12; Gaps 6;

QY 396 KAGKATYDSCNCAACHQPDGKGVNPAFPLAN--SDYLNADHRAASIVA--NGLSGKITV 451
DB 18 QADGAKIVNOCAGCIGNQCGIIPGAFPLAGHVAELIAEGREYILVLVGLGQGLIV 77
QY 452 NGYESVPAIAL-LSDOQIANVITYTTLNSFGN---KGGQ-LSADDAVK--AKTKP 501
DB 78 KGMKVYGVVSSPAQLKDEBIAVLNHIATAMGDAAKVKGFPTAEVKKLPAKKLTP 135

RESULT 17
ID CPA2_PSESIM STANDARD; PRT; 589 AA.
AC P59571;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper resistance protein A homolog precursor.
GN COPA OR PSP703914.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxId=323;
RN [1]
RP SEQUENCE FROM N.A.

```


CC -1- FUNCTION: MAY BE INVOLVED IN A REDOX SYSTEM INVOLVING ASCORBIC
 CC ACID.
 CC -1- CATALYTIC ACTIVITY: 2 L-ascorbate + O(2) = 2 dehydroascorbate + 2
 CC H(2O).
 CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
 CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
 CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: TO CERULOPLASMIN AND LACCASE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X64257; CAA45554.1; -.
 CC PIR: S23763; S23763.
 CC DR InterPro: IPR001117; Cu-oxidase.
 CC Pfam: PF00394; Cu-oxidase; 1.
 CC KW Oxidoreductase; Copper; Glycoprotein; Multigene family; Signal.
 CC FT SIGNAL 1 23
 CC FT CHAIN 1 23
 CC FT METAL 24 535
 CC FT CARBOHYD 444 444
 CC FT CARBOHYD 33 33
 CC FT CARBOHYD 61 61
 CC FT CARBOHYD 110 110
 CC FT CARBOHYD 330 330
 CC FT CARBOHYD 350 350
 CC FT CARBOHYD 422 422
 CC SQ SEQUENCE 555 AA; 62130 MW; 5BP23C9D73EA6C6A CRC64;
 CC
 CC Query Match 5.5%; Score 145; DB 1; Length 555;
 CC Best Local Similarity 26.1%; Pred. No. 0.0091;
 CC Matches 71; Conservative 37; Mismatches 132; Indels 32; Gaps 12;
 CC
 CC QY 100 GVEYGFMTGCGPGCMIRREGDTLEVOFSNHPD-----SKMHNVDFAHATGPGG 152
 CC DB 43 GVPQVILNGQEPGNNSTNNVIVFNLDPEFLITWNGIHRKNCMDGTFPTM 102
 CC QY 153 AEAFTAPGHTSTPSKAL-QEGLVYHCAVAVGSHIANGYGLILVEPKGLP---K 207
 CC DB 103 CP---IMPSTNYTHFPQPDQIGSYFY---PTTGKRAAGGCGRNVSRLLIPVPYAD 156
 CC QY 208 VKKEYVMGDPFTKCKGEQGLQPF-DMEKAI-REDAEVYVFNQSVGLTGENA--LKA 263
 CC DB 157 PEDDYTVLIGDWYTK--SHTOLKFKLGDGRITGRPDG--IYNGKSGKGDSDALEFTL 211
 CC QY 264 KVGETVRLFVNGGAPLTSFVIGEIFDKVHPEGKGSHNNIOTLLIPAG--GAITE 320
 CC DB 212 KKGKTVRAICNVGVTSINFRIOHKKMLVNEGSHVLYONDIDLVDVAGCGFTIYVA 271
 CC QY 321 FRYDVPGDYVLVDHALFRAFNKGAUGILKVEG 352
 CC DB 272 NQ--EPKDYVVAASRFLKTVITTTGLLRVEG 301
 CC
 CC RESULT 21
 CC FAS_BOVIN STANDARD; PRT; 2211 AA.
 CC ID AC Q28107; Q28108;
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Coagulation factor V precursor (Activated protein C cofactor) .
 CC GN F5.
 CC OS Bos taurus (Bovine) .
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Bovinae; Bos.
 CC NCBI_Taxid=9913;
 CC OK

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92147638; PubMed=1737753;
 RA Guinjo E.R., Besson C.T., Mann K.G., Macgillivray R.T.;
 RT "The complete cDNA sequence of bovine coagulation factor V.";
 RL J. Biol. Chem. 267:2971-2978(1992).
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -1- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14
 CC AA repeats.
 CC -1- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).
 CC -1- SIMILARITY: Contains 3 F5/8 type A domains.
 CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M81440; AAA30512.1; -.
 CC DR EMBL: M81441; AAA30513.1; -.
 CC DR PIR: A42580; KPBO5.
 CC DR HSSP: P12259; 1CZ7.
 CC DR InterPro: IPR001117; Cu-oxidase.
 CC DR InterPro: IPR000421; FA58_C.
 CC DR Pfam: PF00394; Cu-oxidase; 3.
 CC DR Pfam: PF00754; F5_F8_type_C; 2.
 CC DR SMART: SM00231; FA58C_2.
 CC DR PROSITE: PS01285; FA58C_1; 2.
 CC DR PROSITE: PS01286; FA58C_2; 2.
 CC DR PROSITE: PS50022; FA58C_3; 2.
 CC DR PROSITE: PS00079; MULTICOPPER OXIDASE1; 2.
 CC KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
 CC Repeat.
 CC FT SIGNAL 1 28
 CC FT CHAIN 29 2211
 CC FT CHAIN 29 741
 CC FT PEPTIDE 742 1564
 CC FT
 CC FT CHAIN 1565 2211
 CC FT DOMAIN 30 327
 CC FT DOMAIN 30 193
 CC FT DOMAIN 203 327
 CC FT DOMAIN 348 686
 CC FT DOMAIN 348 525
 CC FT DOMAIN 535 686
 CC FT DOMAIN 696 1564
 CC FT DOMAIN 1124 1151
 CC FT DOMAIN 1124 1137
 CC FT REPEAT 1138 1151
 CC FT REPEAT 1168 1453
 CC FT REPEAT 1188 1196
 CC FT REPEAT 1197 1205
 CC FT REPEAT 1206 1214
 CC FT REPEAT 1215 1223
 CC FT REPEAT 1224 1232
 CC FT REPEAT 1233 1241
 CC FT REPEAT 1242 1250
 CC FT REPEAT 1251 1259
 CC FT REPEAT 1260 1268
 CC
 CC POTENTIAL.
 CC COAGULATION FACTOR V.
 CC HEAVY CHAIN (BY SIMILARITY).
 CC ACTIVATION PEPTIDE (CONNECTING REGION)
 CC (BY SIMILARITY).
 CC LIGHT CHAIN (BY SIMILARITY).
 CC F5/8 TYPE A 1.
 CC PLASTOCYANIN-LIKE 1.
 CC PLASTOCYANIN-LIKE 2.
 CC F5/8 TYPE A 2.
 CC PLASTOCYANIN-LIKE 3.
 CC PLASTOCYANIN-LIKE 4.
 CC B.
 CC 2 X 14 AA TANDEM REPEATS.
 CC 1-1.
 CC 1-2.
 CC 30 X 9 AA APPROXIMATE TANDEM REPEATS OF
 CC [AS]-L-S-P-D-[LP]-(GS)-Q-(TE).
 CC 2-1.
 CC 2-1.
 CC 2-2.
 CC 2-3.
 CC 2-4.
 CC 2-5.
 CC 2-6.
 CC 2-7.
 CC 2-8.
 CC 2-9.

```

FT REPEAT 1269 1277 2-10.
FT REPEAT 1278 1286 2-11.
FT REPEAT 1287 1295 2-12.
FT REPEAT 1296 1304 2-13.
FT REPEAT 1305 1313 2-14.
FT REPEAT 1314 1322 2-15.
FT REPEAT 1323 1331 2-16.
FT REPEAT 1332 1340 2-17.
FT REPEAT 1341 1349 2-18.
FT REPEAT 1350 1358 2-19.
FT REPEAT 1359 1367 2-20.
FT REPEAT 1368 1376 2-21.
FT REPEAT 1377 1385 2-22.
FT REPEAT 1386 1394 2-23.
FT REPEAT 1395 1403 2-24.
FT REPEAT 1404 1412 2-25.
FT REPEAT 1413 1421 2-26.
FT REPEAT 1422 1430 2-27.
FT REPEAT 1431 1439 2-28.
FT REPEAT 1440 1444 2-29. (PARTIAL).
FT REPEAT 1445 1453 2-30.
FT DOMAIN 1569 1738 F5/8 TYPE A 3.
FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1894 2048 PLASTOCYANIN-LIKE 6.
FT DOMAIN 2053 2208 F5/8 TYPE C 1.
FT SITE 741 742 F5/8 TYPE C 2.
FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1564 1565 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT DISULFID 167 193 PROBABLE.
FT DISULFID 499 525 PROBABLE.
FT DISULFID 1712 1738 PROBABLE.
FT DISULFID 1894 2048 BY SIMILARITY.
FT DISULFID 2053 2208 SULEFATION (POTENTIAL).
FT MOD_RES 697 701 SULEFATION (POTENTIAL).
FT MOD_RES 701 730 SULEFATION (POTENTIAL).
FT MOD_RES 730 730 SULEFATION (POTENTIAL).
FT MOD_RES 1513 1513 SULEFATION (POTENTIAL).
FT MOD_RES 1529 1529 SULEFATION (POTENTIAL).
FT MOD_RES 1537 1537 SULEFATION (POTENTIAL).
FT MOD_RES 1541 1541 SULEFATION (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 902 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 587 592 N-TERMINAL (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2211 AA; 248961 MM; CBBF90B738667C45 CRC64;

```

Query Match 5.5%; Score 143.5; DB 1; Length 2211;
Best Match Similarity 20.8%; Pred. No. 0.07;

```

Matches 80; Conservative 50; Mismatches 133; Indels 121; Gaps 16;
QY 25 NQADKAQPKSTVDAAKTANADNAAGQEGELPVIDATVTHAPEVPPDR----- 78
DB 328 NQAKTRNPKKLTDRQRRHKKMEYFIAE-----EVIDVAFIIPANDKKYRSIH 379
QY 79 -----DHPKVVVVKETVEKVMRLADVEYQFMTF-----GGQVPCQMIRV 119
DB 380 LNFPSNRIGKHKKYKVVYK-----QYQDSDFTKRLIEDPSSGGDILGPILRA 425
QY 120 REGDTIEVQFSNHPD---SKPKNVDF-----HAATGPGGAGASFPAPHTSTFSRK 169
DB 426 QVRDTLKVIFKXMASRSYSIYPHGVTFSPYDNEVSSSTSGSNTMIRAVRPGETTYTKWN 485
QY 170 AL---OPGLVYVHCAVAP-----VGMHIANMGYGLIYVEREGPK---VDKEYVV 214
DB 486 ILESDEPTENDAOCLTRPYNSVDITRDILASGLILCKRSLSDRGIQAAIIEQA 545
QY 215 MQGDF-YTKGKYEGQLOPF--DMEKAIREDAY---VFNQSVGALTGENALAKAYGE 267
DB 546 VPAVFDENKSWYIEDNITKFCENPEKVRDDPKFYESNIMSNFTLPALNG----- 595
QY 268 TVRLFVNGGPNLTSSFVHIGEIFD-----KAFPEGCK---GENHNQ 307
DB 596 -----YVPESIPILGFCFDDTVQWHPGCVGTQNDILTIHTGHSFIYGRHEDT 644
QY 308 TTLIPAGGAITEFKVDVGDVYL 331
DB 645 LTLFPGQESVT-VTMNVGTWML 667

RESULT 22
ID_FAS_PIG STANDARD; PRT; 2258 AA.
AC Q9GEP1;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DE 15-SEP-2003 (rel. 42, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP DOMAINS.
RC TISSUE=Liver;
RX MEDLINE=21121490; PubMed=11229814;
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA Kim H.K.W.;
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
RT protein modeling of membrane binding sites and comparative anatomy of
RT domains."
RL Cell. Mol. Life Sci. 58:148-159 (2001).
CC -1- FUNCTION: Coagulation factor V is a cofactor that participates
CC with factor Xa to activate prothrombin to thrombin.
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -1- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1
CC and C2 may be involved in membrane binding.
CC -1- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -1- SIMILARITY: Contains 3 F5/8 type A domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----
DR EMBL: AF191308; AAG28381.1; -
DR HSSP: P12259; 1CZT.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00754; Cu-oxidase; 3.
DR SMART: SM00231; FA58C_2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS50022; FA58C_3; 2.
DR PROSITE: PS50079; MULTICOPPER_OXIDASE1; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 2258 COAGULATION FACTOR V.
FT CHAIN 23 737 HEAVY CHAIN (BY SIMILARITY).
FT PEPTIDE 738 1611 ACTIVATION PEPTIDE (CONNECTING REGION)
(BY SIMILARITY).
FT CHAIN 1612 2258 LIGHT CHAIN (BY SIMILARITY).
FT DOMAIN 30 329 FS/8 TYPE A 1.
FT DOMAIN 203 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN 348 329 PLASTOCYANIN-LIKE 2.
FT DOMAIN 348 683 FS/8 TYPE A 2.
FT DOMAIN 535 525 PLASTOCYANIN-LIKE 3.
FT DOMAIN 691 1611 PLASTOCYANIN-LIKE 4.
FT DOMAIN 1168 1539 B.
FT REPEAT 1168 1176 41 X 9 AA APPROXIMATE TANDEM REPEATS OF
T-L-S-P-D-L- (GS) - (HQ) - T.
FT REPEAT 1177 1185 1.
FT REPEAT 1186 1194 2.
FT REPEAT 1195 1203 3.
FT REPEAT 1204 1212 4.
FT REPEAT 1213 1221 5.
FT REPEAT 1222 1230 6.
FT REPEAT 1231 1239 7.
FT REPEAT 1240 1248 8.
FT REPEAT 1249 1257 9.
FT REPEAT 1258 1266 10.
FT REPEAT 1267 1275 11.
FT REPEAT 1276 1284 12.
FT REPEAT 1285 1293 13.
FT REPEAT 1294 1302 14.
FT REPEAT 1303 1311 15.
FT REPEAT 1312 1320 16.
FT REPEAT 1321 1329 17.
FT REPEAT 1330 1338 18.
FT REPEAT 1339 1347 19.
FT REPEAT 1348 1356 20.
FT REPEAT 1357 1365 21.
FT REPEAT 1366 1374 22.
FT REPEAT 1375 1383 23.
FT REPEAT 1384 1392 24.
FT REPEAT 1393 1401 25.
FT REPEAT 1402 1410 26.
FT REPEAT 1411 1419 27.
FT REPEAT 1420 1428 28.
FT REPEAT 1429 1437 29.
FT REPEAT 1438 1446 30.
FT REPEAT 1447 1455 31.
FT REPEAT 1456 1464 32.
FT REPEAT 1465 1473 33.
FT REPEAT 1474 1482 34.
FT REPEAT 1483 1491 35.
FT REPEAT 1492 1500 36.
FT REPEAT 1501 1509 37.
FT REPEAT 1510 1518 38.
FT REPEAT 1519 1527 39.
FT REPEAT 1531 1539 40.

FT DOMAIN 1616 1941 FS/8 TYPE A 3.
FT DOMAIN 1616 1785 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1795 1941 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1941 2095 FS/8 TYPE C 1.
FT DOMAIN 2100 2255 FS/8 TYPE C 2.
FT SITE 737 738 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT SITE 1029 1030 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT SITE 1611 1612 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT DISULFID 167 193 PROBABLE.
FT DISULFID 499 525 PROBABLE.
FT DISULFID 1759 1785 PROBABLE.
FT DISULFID 1941 2095 BY SIMILARITY.
FT DISULFID 2100 2255 BY SIMILARITY.
FT MOD_RES 692 692 SULFATION (POTENTIAL).
FT MOD_RES 696 696 SULFATION (POTENTIAL).
FT MOD_RES 724 724 SULFATION (POTENTIAL).
FT MOD_RES 726 726 SULFATION (POTENTIAL).
FT MOD_RES 745 745 SULFATION (POTENTIAL).
FT MOD_RES 1560 1560 SULFATION (POTENTIAL).
FT MOD_RES 1576 1576 SULFATION (POTENTIAL).
FT MOD_RES 1581 1581 SULFATION (POTENTIAL).
FT MOD_RES 1584 1584 SULFATION (POTENTIAL).
FT MOD_RES 1588 1588 SULFATION (POTENTIAL).
FT MOD_RES 1631 1631 SULFATION (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1057 1057 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1066 1066 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1480 1480 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1537 1537 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1597 1597 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1886 1886 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2243 2243 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2258 AA; 256078 MW; 9159B9E0076A2ACC CRC64;

Query Match 5.2%; Score 137; DB 1; Length 2258;
Best Local Similarity 18.4%; Pred. No. 0.21;
Matches 96; Conservative 65; Mismatches 171; Indels 190; Gaps 20;

Qy 25 NOADKAQPKSTVDAAKTAANADNAAQGEGLPVIDATVTHAEVPPVDR----- 78
Db 328 NCAKTRKPKLTLDQRHRIKWEYFLAAE-----EVLWDVAPILIPANMDKRYRLH 379
Qy 79 -----DHPARVVVM-----ETVEKMRLLADGVEXQFWTFGGQVPGQIMRVREG 122
Db 380 LDNFSNQIGKHKKVKKVYQYODESFTKRLBNPNKEDS-----ILGPVIRAGVR 428
Qy 123 DTIEVQFSN-----HPDSKMPHNVDPHATGPGGGAASFTAGHSTSFSL- 171
Db 429 DTLKIVKVMASRSYSIYPHGVTSPYEDDVNASSSTSDNNTMIRAVQGETIYTKWNILE 488

QY 172 --QPGLYVYHCAVAP-----VGMHIANMGYLIVPEKGLPK-----VDKEYVYVMOG 217
 DB 469 SEPEPNDACCLTRPYSVNDITRDIASLIGLLICRSKSDKGIQRTALIEQKAVYA 548
 QY 218 DF-YTKKXGEGQLOPF--DMEKAREDAEYVFNQSGVLTGEMAKAKVETVRLFVG 274
 DB 549 VDEKMSVYIEDNIYKFCENPEKVKDDPKF-----YESNIMSTINGVYP 593
 QY 275 NGCPNLTSEFHYIGELFD-----KVHEEGK---GEMHNIQTLLIPAG 314
 DB 594 ESIFPL-----GFCFDDTVQMHGCSVRTHDNLITHTGHSFYGKHEDTLTFEPR 646
 QY 315 GAITEFKVDVPGDYVLVD-----HAIFPAFNKALG 346
 DB 647 GESVTVMDVNGVTWMLTMSNPRKKQLKFRDYKCIKRDDEDESEIYERSSSTTLT 705
 QY 347 ILKV-----EGEENHEIYSHK-----QTDVAVL 369
 DB 706 TRKHMDSENKEEENDEYDQDLASVLGIRSFRRSSLYQEDDEFNLTALENSEPT 765
 QY 370 PEGARQALITQEARPK---TPAPANL-----QEQIKAG 398
 DB 766 PSTDRADVNSSSSPGNISRAPANTPTETPRKILPHEATKAG 807
 RESULT 23
 LAC1_THACU STANDARD; PRT; 576 AA.
 ID LAC1_THACU STANDARD; PRT; 576 AA.
 AC P56193;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Laccase 1 precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
 GN (Urethol oxidase) (biphenol oxidase).
 OS LCC1.
 OS Thanaethorus cucumeris (Black scurf of potato) (Rhizoctonia solani).
 OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Heterobasidiomycetidae; Ceratobasidiales; Ceratobasidiaceae;
 OC Thanaethorus.
 OC NCBI_TaxID=107832;
 OK (1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=R2;
 RX MEDLINE=96171523; PubMed=8598061;
 RA Wahlström J.A., Xu F., Brown K.M., Brown S.H., Goughly E.J.,
 RA Haliker T., Kauppinen S., Pederson A., Schneider P.;
 RT "The identification and characterization of four laccases from the
 RT plant pathogenic fungus Rhizoctonia solani.";
 RL Curr. Genet. 29:395-403(1996).
 CC - FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE). THIS ISOZYME SHOWS OPTIMAL ACTIVITY AT PH 6.
 CC - CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoemulquinone + 2
 CC H(2)O.
 CC - COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC - SUBUNIT: Homodimer.
 CC - SUBCELLULAR LOCATION: Secreted.
 CC - TISSUE SPECIFICITY: IN MYCELIA, AT A LOWER LEVEL THAN LCC4.
 CC - SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC - SIMILARITY: Contains 3 plastocyanin-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; 254275; -; NOT_ANNOTATED_CDS.
 CC DR PIR; S68117; S68117.
 CC DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR002355; MultiCu oxidase2.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
 KW Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
 KW Lignin degradation; Multigene family.
 FT SIGNAL 19
 FT CHAIN 20 576
 FT DOMAIN 21 145
 FT DOMAIN 157 304
 FT DOMAIN 7 304
 FT METAL 82 82
 FT METAL 84 84
 FT METAL 127 127
 FT METAL 129 129
 FT METAL 471 471
 FT METAL 474 474
 FT METAL 476 476
 FT METAL 523 523
 FT METAL 524 524
 FT METAL 525 525
 FT METAL 529 529
 FT CARBOHYD 41 41
 FT CARBOHYD 182 182
 FT CARBOHYD 228 228
 FT CARBOHYD 294 294
 FT CARBOHYD 368 368
 SQ SEQUENCE 576 AA; 64377 MW; 64377 MW; CSEB32918038AE48 CRC64;
 Query Match 5.1%; Score 133.5; DB 1; Length 576;
 Best Local Similarity 20.2%; Pred. No. 0.061;
 Matches 106; Conservative 63; Mismatches 175; Indels 181; Gaps 26;

QY 86 VMKEVEKMKRLA-----DGVYQFWTFGGVPGQMRVREGDILEVQFSN---HPDSK 136
 DB 17 VLAARVEYGLKTSIDSEIAPDGVKRNATLVNGYPPPLFANQGDTLKVNQKLTNPEWY 76
 QY 137 MDHNVDFHAA-----TGPGGABASFT-----APGHTSIFSKA-LQPLVYVHCAV 182
 DB 77 RTTSHHMGHLLHRNADDGP-----STVVOCPVPRASYTYTIPLDQGTSTYVW 127
 QY 183 AFGVGHIANMGY-LIVPEKE--GLPRVDEKEYVM-OGDFYTKKXGEGQLOPPDMEX 237
 DB 128 SHLSQYVDGLGRLVYDPPHRLVYDVDEKTVLLIGDWY-----HESK 175
 QY 238 AL-----REDAEYVFNQSGVLTGEMAKAKVETVRLFVGNGSPNLT 281
 DB 176 AIIASGNITRQRPVSATNGK-GRDPDPTANPDTLYTLKVRGGRVRLVINSSEIAS 234
 QY 282 SSFHYIGELFDKVFEGKGENHNIQTLLIPAGAAITFEKYDV----- 325
 DB 235 PRFSV-----EGHKV--TVIADGVSTKRYOYADPILAGRIDCVLEAN 277
 QY 326 --PGDYVL-----VDHAFPAFN---KGA 344
 DB 278 QEPDTVYNAIPLTNVKNKTAQALVYEDRRRYPHPKGYRKMSVSEALIKWNHGHKGG 337
 QY 345 LGILKKEG-----EENHEIYS-----HKQTDVAVLEGAPOALDTPAPANL 391
 DB 338 RGLSGHGLKARMEIGSHLSRSVVRKONETTVWDESKLVPLEYGAAGSKPADD 397
 QY 392 QEQIKRATVYDSCNCAHQPDGKGVNAPPL-----ANSDYLNADHARA 438
 DB 398 VLDL-----TFGLNFTAGWMMI-NGIPYSPKIPITLLKLTDEDDGVTEDFTEEH---T 448
 QY 439 SIYANGLSKITVNGQYSVAPALASDQOIANVITYLNSFGN 483
 DB 449 VILPKKCIIEFNKSGSIPITHPVHLGH-----TWVVOFGN 487
 RESULT 24
 LAC4_THACU STANDARD; PRT; 531 AA.
 ID LAC4_THACU STANDARD; PRT; 531 AA.

AC 002081; 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUL-1999 (rel. 38, Last annotation update)
 DE Laccase 4 precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
 DE (urihiol oxidase) (diphenol oxidase).
 GN LCC4.
 OS Thanatephorus cucumeris (Black scurf of potato) (Rhizoctonia solani).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Heterobasidiomycetidae; Ceratobasidiales; Ceratobasidiaceae;
 OC Thanatephorus.
 OC NCBI_Taxid=107832;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=RS22;
 RX MEDLINE=96171523; PubMed=8598061;
 RA Wahlechner J.A., Xu F., Brown K.M., Brown S.H., Golightly E.J.,
 RA Halter T., Kauppinen S., Pederson A., Schneider P.,
 RT "The identification and characterization of four laccases from the
 RT plant pathogenic fungus Rhizoctonia solani."
 RL Curr. Genet. 29:395-403 (1996).
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE). THIS ISOZYME SHOWS OPTIMAL ACTIVITY AT PH 7.
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoemquinone + 2
 CC H(2)O.
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: IN MYCELIA, AT A HIGHER LEVEL THAN LCC1, LCC2
 CC AND LCC3.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; 254277; CAA91042.1; -
 CC PIR; S68120; S68120.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; MulticCu_oxidase2.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER OXIDASE1; FALSE NEG.
 DR PROSITE; PS00080; MULTICOPPER OXIDASE2; FALSE NEG.
 KM Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
 KM Lignin degradation; Multigene family; Polymorphism.
 FT SIGNAL 1 19
 FT CHAIN 20 531
 FT DOMAIN 23 146
 FT DOMAIN 158 315
 FT DOMAIN 384 507
 FT METAL 83 83
 FT METAL 85 85
 FT METAL 128 128
 FT METAL 130 130
 FT METAL 427 427
 FT METAL 430 430
 FT METAL 432 432
 FT METAL 439 439
 FT METAL 479 479
 FT METAL 480 480
 FT METAL 481 481
 FT METAL 484 484
 FT CARBOHYD 66 66
 FT CARBOHYD 109 109
 FT CARBOHYD 186 186
 FT CARBOHYD 231 231
 FT CARBOHYD 280 280

FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 42 42 P -> S.
 FT VARIANT 119 119 H -> R.
 FT VARIANT 246 246 R -> S.
 FT VARIANT 256 256 P -> L.
 FT VARIANT 261 261 P -> A.
 SQ SEQUENCE 531 AA; 57545 MW; 737339803F75AB19 CRC64;
 Query Match 5.0%; Score 131; DB 1; Length 531;
 Best Local Similarity 20.3%; Pred. No. 0.083;
 Matches 93; Conservative 54; Mismatches 203; Indels 108; Gaps 17;
 QY 59 LPVIDAIVTHAEVPEPPVDRHPAKVVK--METVEKYRLADGYEYQFWFGGVPGQM 116
 DB 8 LPLAAVST-----PAPAAYRNKYFQDIKNVNAVDPGQRPVSNGLVPGTL 54
 QY 117 IRRREGDTIEVQFSNH---PDSKMHNVDFH---AATGPGGAASFTAP-GHSTSEF 168
 DB 55 ITANKGDTLRINVTNQLTDPFSKRRATTHMGLFQATTADBDGPAFVTQCPAQNLSYTY 114
 QY 169 KAL---OPGLVYHCAVAPVGMHNGMYG-LIVER---KEGPKYDKEXYVMQGPXY 220
 DB 115 ELPFLHGOTGMWYH---AHLSQYVDGRLGPLVITDVPDPRKSRVDVDASTVWLEBDY 171
 QY 221 -----TKKXGEGG---LOPFMEKAIREDAEYVFNVSVALTGENALAKVGETVR 270
 DB 172 HTPAPVLEKQMFSTNTALLSPDPSGLINGKRVY-----CGPAVPSVIVNKKGRYR 226
 QY 271 LFGVNGGNLTSSFFVIGEIFDKVAFEGGKBNHNIQTLLIPAGAAITFEFVDVPGDYV 330
 DB 227 LRVINASAIGSTFTFISBHRRLVLEADG-----IPHOPLVPDSFOIVAGORYS 274
 QY 331 LVDAHIFAPFKKGAIGLKVEGENHEIYSHKQTDAYVYLPGAPOAIDTQE----- 381
 DB 275 VIVEANQTAANYVIRAPVTWAGAGTNANLDPNNVFAVLHBCAPAEPTTGGSAIGTAL 334
 QY 382 -----APKTPAPANLOEQIKAKATYDS-----NCAACHOPDQKGVNAPPP 423
 DB 335 VEENHLALINGAPGSGAPADVSINLAIGRSTVDGILFTFNINIKYAP-----SLPT 387
 QY 424 LANSDYLNADHARAASIVANGLSKITVNGQYESVMP 461
 DB 388 L-----LKIANNASNDADPTPNEHTIVLP 412
 RESULT 25
 FAS_HUMAN STANDARD; PRT; 2224 AA.
 AC P12259; Q14285;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GN F5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT MET-1764.
 RX MEDLINE=92232668; PubMed=1567832;
 RA Cripe L.D., Moore K.D., Kane W.H.;
 RT "Structure of the gene for human coagulation factor V";
 RL Biochemistry 31:3777-3785 (1992).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT GLU-925.
 RX MEDLINE=87260886; PubMed=3110773;
 RA Jenny R.J., Pittman D.D., Toole J.J., Kitz R.W., Aldape R.A.,
 RA Hewick R.M., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850 (1987).
 RN [3]
 RP SEQUENCE OF 1-1600 FROM N.A., AND VARIANTS GLU-925 AND ILE-1285.

RA MEDLINE=8107560; PubMed=2827731;
RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RT "Cloning of cDNAs coding for the heavy chain region and connecting
RT region of human factor V, a blood coagulation factor with four types
RT of internal repeats.";
RL Biochemistry 26:6508-6514(1987).
[4]
RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX MEDLINE=86313665; PubMed=3092220;
RA Kane W.H., Davie E.W.;
RT "Cloning of a cDNA coding for human factor V, a blood coagulation
RT factor homologous to factor VIII and ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
[5]
RP PARTIAL SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=9303619; PubMed=8454869;
RA Shen N.L., Fan S.-T., Pyatt J., Graff R., Lapolla R.J.,
RA Edgington T.S.;
RT "The serine protease cofactor factor V is synthesized by
RT lymphocytes.";
RL J. Immunol. 150:2992-3001(1993).
[6]
RP SULFATION.
RX MEDLINE=94264012; PubMed=8204629;
RA Pittman D.J., Tomkinson K.N., Michnick D., Seligsohn U.,
RA Kaufman R.J.;
RT "Posttranslational sulfation of factor V is required for efficient
RT thrombin cleavage and activation and for full procoagulant activity.";
RL Biochemistry 33:6952-6959(1994).
[7]
RP SULFATION.
RX MEDLINE=90366699; PubMed=2168225;
RA Horton G.L.;
RT "Sulfation of tyrosine residues in coagulation factor V.";
RL Blood 76:946-952(1990).
[8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
RX MEDLINE=20052169; PubMed=1058686;
RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
RA Fuentes-Prior P.;
RT "Crystal structures of the membrane-binding C2 domain of human
RT coagulation factor V.";
RL Nature 402:434-439(1999).
[9]
RP VARIANT MET-1764.
RX MEDLINE=95179146; PubMed=7874144;
RA Bayson T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;
RT "A polymorphism in the human coagulation factor V gene.";
RL Hum. Mol. Genet. 3:2085-2085(1994).
[10]
RP VARIANT APCR GLN-534.
RX MEDLINE=94217810; PubMed=8164741;
RA Bertina R.M., Koelmaan B.P.C., Koster T., Rosendaal F.R.,
RA Dirven R.J., de Ronde H., van der Velden P.A., Rietma P.H.;
RT "Mutation in blood coagulation factor V associated with resistance to
RT activated protein C.";
RL Nature 369:64-67(1994).
[11]
RP VARIANTS ILE-1285 AND ARG-1327.
RX MEDLINE=96351768; PubMed=8713778;
RA Lunghi B., Iacovello L., Gemmati D., Dilasio M.G., Caseroldi E.,
RA Pinotti M., Castellan G., Redaelli R., Mariani G., Marchetti G.,
RA Bernardi F.;
RT "Detection of new polymorphic markers in the factor V gene:
RT association with factor V levels in plasma.";
RL Thromb. Haemost. 75:45-48(1996).
[12]
RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
RX MEDLINE=98122763; PubMed=9454741;
RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";

RL Blood 91:1135-1139(1998).
[13]
RP VARIANT APCR THR-334.
RX MEDLINE=98122764; PubMed=9454742;
RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
RT resistance to activated protein C.";
RL Blood 91:1140-1144(1998).
[14]
RP VARIANTS HIS-107, THR-413, LYS-513, SER-809, THR-817, ARG-858;
RX ARG-865; GLU-925, GLN-1146, ALA-1530, SER-1685, VAL-1749, MET-1764;
RP ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.
RX MEDLINE=99318093; PubMed=10391209;
RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O.,
RA Lander E.S.;
RT "Characterization of single-nucleotide polymorphisms in coding regions
RT of human genes.";
RL Nat. Genet. 22:231-238(1999).
[15]
RP ERRATUM.
RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.O.,
RA Lander E.S.;
RL Nat. Genet. 23:373-373(1999).
CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
CC REPEATS.
CC -1- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
CC -1- DISEASE: Defects in F5 are the cause of Owren parahemophilia
CC (MIM:227400), an hemorrhagic diathesis.
CC -1- DISEASE: Defects in F5 are the cause of resistance to activated
CC protein C (APCR) [MIM:188055], a form of thrombophilia. The APCR
CC mutation is found in about 5% of the population which suggest that
CC a slight thrombotic tendency may confer some advantage in fetal
CC implantation.
CC -1- SIMILARITY: Contains 3 F5/8 type A domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L32779; AAB59401.1; -;
DR EMBL; L32755; AAB59401.1; JOINED.
DR EMBL; L32756; AAB59401.1; JOINED.
DR EMBL; L32757; AAB59401.1; JOINED.
DR EMBL; L32758; AAB59401.1; JOINED.
DR EMBL; L32759; AAB59401.1; JOINED.
DR EMBL; L32760; AAB59401.1; JOINED.
DR EMBL; L32761; AAB59401.1; JOINED.
DR EMBL; L32762; AAB59401.1; JOINED.
DR EMBL; L32763; AAB59401.1; JOINED.
DR EMBL; L32764; AAB59401.1; JOINED.
DR EMBL; L32765; AAB59401.1; JOINED.
DR EMBL; L32766; AAB59401.1; JOINED.
DR EMBL; L32767; AAB59401.1; JOINED.
DR EMBL; L32768; AAB59401.1; JOINED.

DR EMBL: U32769; AAB59401.1; JOINED.
DR EMBL: U32770; AAB59401.1; JOINED.
DR EMBL: U32771; AAB59401.1; JOINED.
DR EMBL: U32772; AAB59401.1; JOINED.
DR EMBL: U32773; AAB59401.1; JOINED.
DR EMBL: U32774; AAB59401.1; JOINED.
DR EMBL: U32775; AAB59401.1; JOINED.
DR EMBL: U32776; AAB59401.1; JOINED.
DR EMBL: U32777; AAB59401.1; JOINED.
DR EMBL: U32778; AAB59401.1; JOINED.
DR EMBL: M16967; AAB52424.1; -
DR EMBL: M16335; AAB59532.1; -
DR PDB: 1C2S; 26-NOV-99.
DR PDB: 1C2T; 26-NOV-99.
DR PDB: 1C2V; 26-NOV-99.
DR PDB: 1FV4; 17-JAN-01.
DR Genew; HGNC:3542; F5.
DR MIM: 227400; -
DR MIM: 134400; -
DR MIM: 188055; -
DR MIM: 227310; -
DR GO: GO:0003801; F: blood coagulation factor activity; TAS.
DR GO: GO:0007596; P: blood coagulation; TAS.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00394; Cu-oxidase; 3.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.

Query Match 4.9%; Score 128.5; DB 1; Length 2224;
Best Local Similarity 19.7%; Pred. No. 0.8;
Matches 82; Conservative 61; Mismatches 150; Indels 123; Gaps 20;

QY 63 DAIVTHAPEPPVPRDHPAKVVMKMT-----VEKVMRLADQVEYQFTPGGQ----- 111
DB 358 EYIMYAPVPIIPAMDKKRSQHLDNFSNIGIKHYKKVMY-----TQDESEFTKHTVNNM 413
QY 112 ---VPGQWIRVREGDTLEVQFSN---HPDSKMFHVDFHAATGPGGAASFTAPGHTS 164
DB 414 KEDGILGPITRAQVDTLTKIVFKMNASRPSYISYPHGVTSPYEDS---VNSSFTS-GRNN 469
QY 165 TFSFALQPG-LYV-----HCAVAP-----VGMHANGMTGLIVEPK 202
DB 470 TM-IAAVPGETTYTKWNLLEFDEPTENDACQITRPSYDVIMIDISGLIGLLICKS 528
QY 203 EGLPK-----VDKEYVWQGF-YTKGKYGEQGLQPF--DMEKAIREDAEVIVNGSNG 253
DB 529 RSLDRRGIGRAADIEQAVFAVFDEKSKWYLEDNINKFCENDEVRDDPKF----- 580
QY 254 ALTGENALKAKVGETVRLFEVNGGPNLTSSFFHVGIEIFD-----KVHF 296
DB 581 -----YESNIMSTINGV-----PESITTLGFCPDYVQWHFCSVGTONEILTIHF 626
QY 297 BGCK--GENNITQTTLLPAGGAITEFKVDVPGDYLVDHAIFFAENKALGILKEVE 353
DB 627 TGHSEFYKRRHEDTLTLFPMRGESVT-----VTMD-----NVGTWMLTSMNS 669
QY 354 ENHEIYSHKQTDVAVLPEGAIPAIDQEAPKTPAPA-----NLOEQIKAGATYD 403
DB 670 PRSKQLRLKFRVVKCIPIDDEDSYEIPEPESTVWATRMHRLPEDESESDADYD 725

RESULT 26
CUBO_YERPE STANDARD; PRT; 533 AA.
AC Q8ZBK0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Blue copper oxidase cubo precursor (Copper efflux oxidase).
GN CUBO OR YPO3409 OR Y0777.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / Biovar Orientalis;
RX MEDLINE=21770413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sepahia M., James K.D., Churche C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston F.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM";
RL J. Bacteriol. 184:4601-4611(2002).
CC -I- FUNCTION: Probably involved in periplasmic detoxification of
CC copper by oxidizing Cu(I) to Cu(II) and thus preventing its uptake
CC into the cytoplasm. Possesses phenoloxidase and ferroxidase
CC activities and might be involved in the production of polyphenolic
CC compounds and the prevention of oxidative damage in the periplasm
CC (By similarity).
CC -I- COFACTOR: This protein belongs to the multicopper oxidases which
CC contain three distinct Cu centers known as type 1 or blue, type 2
CC or normal, and type 3 or coupled binuclear (By similarity).
CC -I- SUBUNIT: Monomer (Probable).
CC -I- SUBCELLULAR LOCATION: Periplasmic. It is exported via the Tat
CC pathway (By similarity).
CC -I- INDUCTION: By Cu²⁺, at increased levels of cytoplasmic cuprous
CC ions (Probable).
CC -I- DOMAIN: The methionine-rich domain could provide binding sites for
CC exogenous copper ions. This methionine-rich region is probably
CC important for copper tolerance in bacteria.
CC -I- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -I- SIMILARITY: Contains 3 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ414157; CAC92639.1; -
CC EMBL: AE013680; AAM84364.1; ALT_INIT.
CC PIR: AC0414; AC0414.
CC InterPro: IPR001117; Cu-oxidase.
CC InterPro: IPR002355; Multicu_oxidase2.
CC InterPro: IPR006311; Tat.
CC Pfam: PF00394; Cu-oxidase; 1.
CC TIGRFAMs: TIGR01409; TAT signal seq; 1.
CC PROSITE: PS00080; MULTICOPPER OXIDASE2; 1.
CC Oxidoreductase; Copper; Metal-binding; Periplasmic; Signal; Repeat;
KM Complete proteome.
KW SIGNAL
FT 1 28
FT CHAIN 29 533 BY SIMILARITY.
FT DOMAIN 68 164 BLUE COPPER OXIDASE CUBO.
FT DOMAIN 165 425 PLASTOCYANIN-LIKE 1.
FT DOMAIN 426 533 PLASTOCYANIN-LIKE 2.
FT DOMAIN 356 415 PLASTOCYANIN-LIKE 3.
FT METAL 102 102 MET-RICH.
FT METAL 104 102 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 104 104 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 142 142 COPPER (TYPE 3) (BY SIMILARITY)
 FT METAL 144 144 COPPER (TYPE 3) (BY SIMILARITY)
 FT METAL 458 458 COPPER (TYPE 2) (BY SIMILARITY)
 FT METAL 461 461 COPPER (TYPE 2) (BY SIMILARITY)
 FT METAL 463 463 COPPER (TYPE 3) (BY SIMILARITY)
 FT METAL 514 514 COPPER (TYPE 3) (BY SIMILARITY)
 FT METAL 515 515 COPPER (TYPE 1) (BY SIMILARITY)
 FT METAL 516 516 COPPER (TYPE 3) (BY SIMILARITY)
 FT METAL 520 520 COPPER (TYPE 1) (BY SIMILARITY)
 FT METAL 525 525 COPPER (TYPE 1) (BY SIMILARITY)
 SQ SEQUENCE 533 AA; 58328 MW; ED1570C9E9B1C135 CRC64;
 Query Match 4.7%; Score 124; DB 1; Length 533;
 Best Local Similarity 22.1%; Pred. No. 0.26;
 Matches 64; Conservative 51; Mismatches 131; Indels 44; Gaps 14;
 QY 72 VPPVDRDHPAKVNVKMEYKVRMLADGVEYQFWTFGGVPGQMIRREGDTIEVOFSN 131
 DB 36 IEPPLQDPANGLNINLQIGSVVW--LPSATQTMWYNGNLGPAIRLQKAVTIDITN 93
 QY 132 H-PDSMPHNVDFHAATGFG--GGAASFTAGHTSTSPKALQFG--LYYHCAVAEV 185
 DB 94 ALPEATVH--WHGLEIPEGVDSGQA-LIQGAKROVTFAVEOPATCWPHPTHSTKT 149
 QY 186 GMIHANGWGLIIVERKEG---LPRK---VDKEYVMQDFFYKYGEGLOPFPMEKA 238
 DB 150 GHQVANGGLGLVILIDSDSETLPLPKQMGVDDLPVILQDKL--DHGOVDIQLDMVTA 207
 QY 239 IREDAEYVFNNG-----SVGALTGENA---LKAQVGETVRLFVNGNG--- 277
 DB 208 VGMFGGRMLTNGVYPOQITPRGVRRLRLNGNANSLNALSDGPMYVIASDGLLAE 267
 QY 278 PNLTSFFHYI-GEIFPKVHEGSGKENHNIQTLLPAGCAATIEFKVNDP 326
 DB 268 PUVVRELPIIMGERFE-VIVDTRDGSLDLVTLFVQMGTLAPFDQPLP 316
 RESULT 27
 LAC3_THACU STANDARD; PRT; 572 AA.
 AC 002079;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Laccase 3 precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
 GN (Urishiol oxidase) (Diphenol oxidase).
 OS Thanatephorus cucumeris (Black scurf of potato) (Rhizoctonia solani).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Heterobasidiomycetidae; Ceratobasidiales; Ceratobasidiaceae;
 OC Thanatephorus.
 OX NCBI_TaxID=107832;
 RP [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=RS2;
 RX MEDLINE=66171523; PubMed=8598061.
 RA Walchliener J.A., Xu F., Brown K.M., Brown S.H., Golightly E.J.,
 RA Halckier T., Kauppinen S., Pederson A., Schneider P.;
 RA "The identification and characterization of four laccases from the
 RA plant pathogenic fungus Rhizoctonia solani.";
 RL Curr. Genet. 29:395-403 (1996).
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzenesemiquinone + 2
 CC H(2)O.
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: IN MYCELA, AT A LOWER LEVEL THAN LCC4.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; 254215; CA90942.1; -
 DR PIR; S68119; S68119.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; Multicu_oxidase2.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
 DR PROSITE; PS00080; MULTICOPPER OXIDASE2; FALSE_NEG.
 KW Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
 KW Lignin degradation; Multigene family; Polymorphism.
 FT SIGNAL 1 18
 FT CHAIN 19 572
 FT DOMAIN 21 145
 FT DOMAIN 157 304
 FT DOMAIN 422 540
 FT METAL 82 82
 FT METAL 84 84
 FT METAL 127 127
 FT METAL 129 129
 FT METAL 470 470
 FT METAL 473 473
 FT METAL 475 475
 FT METAL 522 522
 FT METAL 523 523
 FT METAL 524 524
 FT METAL 528 528
 FT CARBOHYD 162 182
 FT CARBOHYD 228 228
 FT CARBOHYD 294 294
 FT CARBOHYD 367 367
 FT CARBOHYD 405 405
 FT VARIANT 159 159
 FT VARIANT 359 359
 FT VARIANT 418 418
 FT VARIANT 448 448
 SQ SEQUENCE 572 AA; 63747 MW; DF1E2AF9F108CE4 CRC64;
 Query Match 4.7%; Score 124; DB 1; Length 572;
 Best Local Similarity 21.4%; Pred. No. 0.28;
 Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;
 QY 86 VKMEYKVRMLA-----DGEYQFWTFGGVPGQMIRREGDTIEVOFSN---HPDSK 136
 DB 17 VLARIVEYVULKISNGKIPADGVERDALTVMGYPPLIFANKGDLTKVAVQKLNPDWY 76
 QY 137 MBHNVDFHAA-----TGPGGAASFTAGHTSTSPK-ALQPGLYVYHCAVAVGM 187
 DB 77 RTTSHHGLQHRNADDGPAFTQCP-VPQASVYTMPLDGTGTWYH---SHLS 132
 QY 188 HANGWYGLIIVERKEGPKV---DKEYVMQDFFY-TKKR---YQEGLOPFPDMK 237
 DB 133 QYVDGIRGLVLYDPDPHRLRYDIDDEKTVLIDWHTSSKAILANTGNITLQOPDS-- 190
 QY 238 AIREDAEYVFNNGVGLTGEN-----ALKAQVGETVRLFVNGGPNLTSFFHYV 288
 DB 191 -----ATNGK-GRPPDVTTPANPTLTKYKRGKRYLRLVINS--SALASF----- 235
 QY 289 EIFDKVHEGSGKENHNIQTLLIPAGCAATIEFKVNDPVDVYLVDHAIKRAFNKALGIL 348
 DB 236 ---RMSIGHK-----MTVIADGVSTKRYQVD----- 260
 QY 349 KYEGENHEIYSHKQTDVAVLEBGAPOAIDTOEAPCT---PAP-ANLOEOIKAGKATYDS 404
 DB 261 -----SPDILAGRIDAV-----VEANQEPDTWINAPLTVNKAQALIIYED 305

QY 405 NCACHQPDG 414
 Db 306 DRRPYHPKG 315

RESULT 28
 FET5 YEAST
 ID FET5 YEAST STANDARD; PRT; 622 AA.
 AC P43561;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Iron transport multicopper oxidase FET5 precursor (EC 1.11.1.1).
 GN FET5 OR YFL041W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae.";
 RL Nat. Genet. 10:261-268(1995).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=98075360; PubMed=9413439;
 RA Spitzo T., Byersdorfer C., Duesterhoeft S., Eide D.;
 RT "The yeast FET5 gene encodes a FET3-related multicopper oxidase
 RT implicated in iron transport.";
 RL Mol. Genet. 256:547-556(1997).
 CC -1- FUNCTION: IRON TRANSPORT MULTICOPPER OXIDASE, WHICH IS REQUIRED
 CC FOR FERROUS IRON HIGH AFFINITY UPTAKE. MAY BE REQUIRED TO OXIDIZE
 CC FE(II) AND RELEASE IT FROM THE TRANSPORTER. ESSENTIAL COMPONENT OF
 CC COPPER-DEPENDENT IRON TRANSPORT.
 CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
 CC CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
 CC 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE MULTICOPPER OXIDASE FAMILY.
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D50617, BA009199.1; .
 DR PIR: S56214; S56214.
 DR HSSP: P37064; IASO.
 DR SGD: S0001853; FET5.
 DR GO: GO:0004323; F:multicopper ferroxidase iron transport medi. . .; IGI.
 DR GO: GO:0006826; P:iron ion transport; IMP.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; MultiCu_oxidase2.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR PROSITE: PS00079; MULTICOPPER OXIDASE1; 1.
 DR PROSITE: PS00080; MULTICOPPER OXIDASE2; 1.
 KW Glycoprotein; Transmembrane; Signal; Transport; Oxidoreductase;
 KW Copper; Repeat; Metal-binding; Iron transport.
 FT CHAIN 1 18
 FT SIGNAL 1 18
 FT CHAIN 19 622
 FT DOMAIN 19 573
 FT TRANSMEM 574 594
 FT DOMAIN 595 622
 FT DOMAIN 43 146
 FT DOMAIN 192 301
 FT DOMAIN 192 301
 FT PLASTOCYANIN-LIKE 2.

FT DOMAIN 392 514 PLASTOCYANIN-LIKE 3.
 FT METAL 79 79 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 81 81 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 128 128 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 130 130 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 418 418 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 421 421 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 423 423 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 496 496 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 497 497 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 498 498 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 502 502 COPPER (TYPE 1) (BY SIMILARITY).
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 622 AA; 70880 MW; 961757A6C86B7EAF CRC64;

Query Match 4.5%; Score 118; DB 1; Length 622;
 Best Local Similarity 20.9%; Pred. No. 0.83;
 Matches 93; Conservative 54; Mismatches 171; Indels 126; Gaps 22;

QY 99 DGV-EYGTWTFGGGPGMIRVREGDTIEVQPSNHPDSKMPHNVFHA----- 145
 Db 35 DGLHEKRNIGNGBEPPLDPIHVEKGRVETLYTNGFQDNTATSLFHLFQNTSLGNDLO 94
 QY 146 ATGPGGAGASFTAPGHTSTESFKA-LQPGYVYHCAVAPGMHANGMYGLIIVEPYEG 204
 Db 95 MDGPMVYQCP1-VPGQYLYNFTVPEVGVFWNH--AHNGAQQGDGMRGAFIHDEE 150
 QY 205 LPRVDKEYVYVNGQDFY-----TK--GKYGEGLOPFDEKALREDAEVVNGSVGA 254
 Db 151 PFEYDHERVITLSPHYHENYKTVTEFSLRNPFGAEPI-----PQNLFPNNTMV 201
 QY 255 LTGNAALAKKQGT-----VRLFVGN-----GGPNLTSSHVIGE 289
 Db 202 ----TLDTFGEYTLFRELNVGLFVSOYILLEHMSIVEVDGYVRNPF----- 248
 QY 290 IFDKVHEFGKGEHNHIDTTLIPAGAAITFEKVDPGDVYLVDHAIPEAFKAGILK 349
 Db 249 --DSIYLSAGR-----MSVLIKAKDKMPTR-----NYAMQIMDETLDVVP 289
 QY 350 VEGEENHEI---YHKQTDVAVLPFGAPQALDTGEAPKTPAPANLQEQIKAG-KATYDSN 405
 Db 290 PELQNTQITQMRVGH-----LPEARALNIEDCDLDRATNDFLEPLIERDLAHYD-- 341
 QY 406 CAACHQ-----PDGKGVNAPFPPLANSYLNADHRAASIVANGSGKITVNGNOYE 457
 Db 342 ---HQIVMDVRMVLGDGVYAF--FNNITVYVPKVPETLTLLT---SGKLASDPRIYG 392
 QY 458 SVMPAIALSDQIANVITYTNSF 481
 Db 393 DNINAQLKNDHDIIEV---LNNY 413

RESULT 29
 LAC1 AGABI
 ID LAC1 AGABI STANDARD; PRT; 520 AA.
 AC O12541;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Laccase I precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
 GN (Utrishol oxidase) (Diphenol oxidase).
 DN LCC1
 OS Agaricus bisporus (Common mushroom).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Agaricaceae; Agaricus.

NCBI_TaxID=5341;
 OX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RN STRAIN=D649; TISSUE=Myxcelium;
 RC MEDLINE=93367392; PubMed=8360614;
 RX Perry C.R., Smith M., Britnell C.H., Wood D.A., Thurston C.F.;
 RT "Identification of two laccase genes in the cultivated mushroom
 Agaricus bisporus";
 RL J. Gen. Microbiol. 139:1209-1218 (1993).
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoquinone + 2
 CC H(2)O.
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, L10664; AAC18877.1; -.
 CC DR HSSP; P37064; LAO2.
 CC DR InterPro: IPR001117; Cu-oxidase.
 CC DR InterPro: IPR002355; MultiCu_oxidase2.
 CC DR Pfam: PF00394; Cu-oxidase; 3.
 CC DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.
 CC DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
 CC KM Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 CC Glycoprotein; Repeat; Multigene family.
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 520
 CC FT DOMAIN 21 145
 CC FT DOMAIN 157 305
 CC FT DOMAIN 375 488
 CC FT METAL 82 82
 CC FT METAL 84 84
 CC FT METAL 127 127
 CC FT METAL 129 129
 CC FT METAL 417 417
 CC FT METAL 420 420
 CC FT METAL 422 422
 CC FT METAL 422 422
 CC FT METAL 470 470
 CC FT METAL 471 471
 CC FT METAL 472 472
 CC FT METAL 476 476
 CC FT CARBOHYD 108 108
 CC FT CARBOHYD 239 239
 CC FT CARBOHYD 299 299
 CC FT CARBOHYD 454 454
 CC FT CARBOHYD 492 492
 CC SQ SEQUENCE 520 AA; 58099 MW; 943DB3F23297B891 CRC64;
 Query Match 4.4%; Score 116; DB 1; Length 520;
 Best Local Similarity 26.6%; Pred. No. 0.91; Indels 42; Gaps 13;
 Matches 59; Conservative 28; Mismatches 93;

DB 130 LSTGYCDGLRGAFFIYVDRDPLRHLXYDVDDSTVITLAEMHYLLAPD--ATNEFFSGII 187
 QY 222 PFDMEKALREDAEYVNGSVGALGEMALAKVETRELV 273
 DB 188 PQVDGSLINGKGR---FNG--GPLTFAVVVVRKRYRLRV 224
 RESULT 30
 NOSZ PARDE STANDARD; PRT; 652 AA.
 ID NOSZ PARDE
 AC Q51705;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nitrous-oxide reductase precursor (EC 1.7.99.6) (N(2)OR).
 GN NOSZ.
 OS Paracoccus denitrificans.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Paracoccus.
 OX NCBI_TaxID=266;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NL188944;
 RX MEDLINE=94062841; PubMed=8243476;
 RA Hoeren F.U., Berke B.C., Ferguson S.J., McCarthy J.E.G.;
 RT "Sequence and expression of the gene encoding the respiratory
 RT nitrous-oxide reductase from Paracoccus denitrificans. New and
 RT conserved structural and regulatory motifs";
 RL Eur. J. Biochem. 218:49-57 (1993).
 CC -1- FUNCTION: NITROUS-OXIDE REDUCTASE IS PART OF A BACTERIAL
 CC RESPIRATORY SYSTEM WHICH IS ACTIVATED UNDER ANAEROBIC
 CC CONDITIONS IN THE PRESENCE OF NITRATE OR NITROUS OXIDE.
 CC -1- CATALYTIC ACTIVITY: N(2) + H(2)O + acceptor = nitrous oxide +
 CC reduced acceptor.
 CC -1- COFACTOR: COPPER (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER; EACH SUBUNIT CONTAINS 2 DINUCLEAR COPPER
 CC CENTERS A AND Z. Z IS THOUGHT TO BE THE SITE OF NITROUS OXIDE
 CC REDUCTION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, TO MITOCHONDRIAL AND
 CC BACTERIAL COX2 SUBUNTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X74792; CAA52798.1; -.
 CC DR PIR; S39409; S39409.
 CC KM Signal; Oxidoreductase; Copper; Metal-binding; Periplasmic.
 CC FT SIGNAL 1 57
 CC FT CHAIN 58 652
 CC FT METAL 595 595
 CC FT METAL 630 630
 CC FT METAL 630 630
 CC FT METAL 632 632
 CC FT METAL 634 634
 CC FT METAL 634 634
 CC FT METAL 638 638
 CC FT METAL 641 641
 CC SQ SEQUENCE 652 AA; 71413 MW; 40492A4FDEDEDA8 CRC64;
 Query Match 4.4%; Score 116; DB 1; Length 652;
 Best Local Similarity 27.8%; Pred. No. 1.2;
 Matches 45; Conservative 18; Mismatches 63; Indels 36; Gaps 7;
 DB 63 DAIVTAPEVPPV---DHDPAKVAVVMEYKVRADGVGYOFTFGGVPQGMTR 118
 QY 505 DALVAH-PSILSDIKSVWRNDP-----MMAETRAQAEADGVDDINMTVEVIRDNKVR 557

QY 119 -----VREGDTIEVQFSNHPD-SKPHNVDFHAATGPGGAEASTAPG 161
 DB 558 VMSSVAPSPSIESFTVEKEDEVTAVITMIDEIDDLTHGFTM-----GNYGVAMIGIQ 611
 QY 162 HSTFSFKALQGLVYVHCANAVPGMHINGMVGLIVPEKE 203
 DB 612 MTSSVTFVANPGVYVYCCWCFCHALHME--MGRMLVPEKE 651

RESULT 31

NTP3_TOBAC STANDARD: PRT; 554 AA.

AC P29162;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Pollen-specific protein NTP303 precursor.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Petit Havana SRI; TISSUE=Pollen;
 RC MEDLINE=92288302; PubMed=1600146;

RA Weterings K., Reijnen W., van Aarsen R., Kortstee A., Spijkers J.,
 RA van Herpen M., Schrauwen J., Mullens G.,
 RT "Characterization of a pollen-specific cDNA clone from Nicotiana
 RT tabacum expressed during microgametogenesis and germination.";
 RL Plant Mol. Biol. 18:1101-1111(1992).

CC -1- FUNCTION: PRECISE FUNCTION AND/OR UNKNOWN BUT PROBABLY HAS AN IMPORTANT
 CC -1- ROLE DURING GERMINATION AND/OR TUBE GROWTH.
 CC -1- SUBCELLULAR LOCATION: Extracellular (Probable).
 CC -1- TISSUE SPECIFICITY: POLLEN.
 CC -1- DEVELOPMENTAL STAGE: APPEAR AFTER THE FIRST HAPLOID MITOSIS AND
 CC ARE EXPRESSED DURING MICROGAMETOGENESIS, GERMINATION AND TUBE
 CC GROWTH.

CC -1- INDUCTION: EXPRESSION REGULATED BY THE HAPLOID GAMETOPHYTE ITSELF.
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, X61146; CAA3454.1; -
 DR PIR, S22495; S22495.
 DR InterPro: IPR001117; Cu-oxidase.
 DR Pfam: PF00394; Cu-oxidase; 1.

KW Signal; Germination; Glycoprotein; Repeat.
 FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 554 POLLEN-SPECIFIC PROTEIN NTP303.
 FT DOMAIN 22 143 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 196 286 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 411 521 PLASTOCYANIN-LIKE 3.

FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 554 AA; 62033 MW; 9D38DAB1F52E2F85 CRC64;

Query Match Best Local Similarity 4.2%; Score 111.5; DB 1; Length 554;
 Matches 91; Conservative 43; Mismatches 160; Indels 143; Gaps 22;

QY 100 GVEYQFWTFGGVPGOMIRVREGDTIEVQFSNHPD-----SKPHNVDFHAATGPGG 152

DB 41 GVPQGIILNGQFPBPRLNCTNNNI VNVFNNLDEPFLFTWNGVQHRKNSVQDTPGTM 100
 QY 153 AASAFAPAGHTSTFSPKRL-OPGLVYVHCANAVPGMHINGMVGLIVPEKEGLP-----K 207
 DB 101 CP---IMPGQNTTFRVQYDQIGSYSPPTTA---LHRAGGIGALNVSRRLIPVPPFN 154
 QY 208 VKEEYVVMQDFFYTKGKGEQGLQPFDMKAIREDAEVYVFNQ--SVGALTG--ENALKA 263
 DB 155 PADENNVFVGDMYNNQ-----HKTLKK-----ILDGRTIGRPDGIINGKSA 197
 QY 264 KVGETVR-LFVNGGPNLTSSF-----HVICEIFDKVHF 296
 DB 198 KVGAEKPELFTWEAKTYRFRFCNLMSSVNI RFQGHMKLVELEGSHTVQNIYDSLQ 257
 QY 297 ECGKEBNNIQITLLPAGGAATFERKVDVPGGVYVDNAIFPAFKGALGILKVEEENH 356
 DB 258 HVGQ---CLSVLTAD-----QEPKDYVLV--VSSRFLKQALSSVAL----- 294
 QY 357 EYSHKQTDVAVLPGAPQALDTPQAPKTPAPANQ-----EQIKAGKATVDSNCAQH 410
 DB 295 -----IRYANGKGA--SPELP--TPPEENTEGIAMSNQRRSPFNNLTASAR-- 339
 QY 411 QPDGKGVNAPPPPLANSDYLNADHAPASIVANGISKITVNGNOYESVMPAIALSDQI 470
 DB 340 -----FN---PQGSYHVGQINIRITIKIF-NSMS-----QV, 366
 QY 471 ANVIYTLNFPNKGQ 487
 DB 367 GGLKRYGLNGISHTNGE 383

RESULT 32

LACS_TRAVI STANDARD: PRT; 527 AA.

AC Q99056;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Laccase 5 precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
 DE (Urishiol oxidase).
 GN LACS.

OS Trimeetes villosa (white-rot fungus).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Trimeetes. . .

OX NCBI_TaxID=47662;
 RN [1]

RP SEQUENCE FROM N.A.
 RP MEDLINE=97128774; PubMed=8973314;

RA Yaver D.S., Golightly E.J.;
 RT "Cloning and characterization of three laccase genes from the
 RT white-rot basidiomycete Trimeetes villosa: genomic organization of the
 RT laccase gene family.";
 RL Gene 181:95-102(1996).
 RP [2]

RP REVISIONS.

RA Yaver D.S., Golightly E.J.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzenesemiquinone + 2
 CC H(2)O.

CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).

CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: Contains 3 plastocyanin-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)

CC EMBL, L78078; AAB47735.2; -

DR PIR: JCS357; JCS357.

DR HSSP: P37064; 1A02.

DR InterPro: IPR001117; Cu-oxidase.

DR InterPro: IPR002355; MultiCu-oxidase2.

DR Pfam: PF00394; Cu-oxidase; 3.

DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.

DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; FALSE NEG.

KW Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
 KW Lignin degradation; Multigene family.

KW SIGNAL 1 23
 FT CHAIN 24 527
 FT DOMAIN 25 150
 FT DOMAIN 162 306
 FT DOMAIN 373 498
 FT METAL 87 87
 FT METAL 89 89
 FT METAL 132 132
 FT METAL 134 134
 FT METAL 425 425
 FT METAL 428 428
 FT METAL 430 430
 FT METAL 480 480
 FT METAL 481 481
 FT METAL 482 482
 FT METAL 486 486
 FT CARBOHYD 74 74
 FT CARBOHYD 77 77
 FT CARBOHYD 156 156
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 242 242
 FT CARBOHYD 276 276
 FT CARBOHYD 317 317
 FT CARBOHYD 358 358
 FT CARBOHYD 366 366
 FT CARBOHYD 393 393
 FT CARBOHYD 402 402
 FT CARBOHYD 464 464
 FT SEQUENCE 527 AA; 56247 MW; F1B638D65FFA478 CRC64;

Query Match 4.2%; Score 111; DB 1; Length 527;
 Best Local Similarity 19.5%; Pred. No. 2.1;
 Matches 101; Conservative 63; Mismatches 187; Indels 168; Gaps 23;

QY 60 PVTDAIVTAPEVPPPPDRDHPAKVNVKMEVKRLADGVEYQFTFGGQVPGCMIRV 119
 DB 27 PVTDLTISNADVP-----DGIIRAVALAGVPPGPIITG 61

QY 120 REGDTLEVG-FSNHPSKMPHNVDF-----ATGPGGGAASFT-----APGHTSTFS 167
 DB 62 NKGDPEQIIVINDLMTETMLKSTTIHWGIFQAGTKWADG--AAVVOGCIATGNSLYID 119

QY 168 FKAL-OPGLVYVCAVAPVGMHIANGVGLILV-----EPKEGLPKYDKEXYVNO-GDFY- 220
 DB 120 FTVPPDQAGTFWYH---SHLSTQYCDGLRGLPLVYVDPDPNALSLYDVDDTTVTLADWYH 176

QY 221 TKRGVEGQLQPPDMEKALREDAEYVFNQSGVGLTGENALKKAKVETVRLPVGNGCPNL 280
 DB 177 TAAKLG-----PAPFAGPSVYLING-LGRFSGD-----GGGATNL 210

QY 281 T-----SSFVIGEIPDKVFEKGKGNHNIOTLLPAGGAAT 318
 DB 211 TVITVTGKKRYRRLVLSICDPNFTSIDGNHMTIIEVDGVNHDALVDVSIQIFAGQRY- 269

QY 319 TEKKVDVPGDYVL-----VDHAIFFRAE-NKGAL-----GLIKVGEENHEIYSHKOT 364
 DB 270 -----SFLNANGSIDNWYIRALPNTGTTDTTGCVNSALIRYVTADIEPTTWTAT 320

QY 365 DAVYLPAGAPQAIPTQZAPKTP-----APANLOE 393
 DB 321 SVIFPLTEIDVLPDNPAPAPGPQVGVGLANSLODFNGSNFFINNTEFVPTVLL-- 378

QY 394 QIKAG--KATYDSNCAACHOPDGKGVNAPFPPLANSYLYNADHARAASIYANGISGKIT 450
 DB 379 QILSGAQDAASLLPNGSVYTLPSNSTIEISFPIITTDGVLNAP-----GAPHPER 428

QY 451 VNGNYESVMPALIASDQIANVITYTLNSRGNKGQLS 489
 DB 429 LHGHTF-SVVSAGSSTFNYPNVPVRDVTYVTSNGSDNVT 466

RESULT 33
 GUNC_CELF1
 ID GUNC_CELF1 STANDARD; PRT; 1101 AA.
 AC P14090;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase C)
 DE (Cellulase C).
 GN CENC.
 OS Cellulomonas fimi.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_Taxid=1708;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 33-42.
 RC STRAIN=ATCC 484;
 RX MEDLINE=92065819; PubMed=1956299;
 RA Coutinho J.B., Moser B., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;
 RT "Nucleotide sequence of the endoglucanase C gene (cenc) of
 RT Cellulomonas fimi, its high-level expression in Escherichia coli, and
 RT characterization of its products.";
 RL Mol. Microbiol. 5:1221-1233(1991).
 RN [2]
 RP SEQUENCE OF 1-64 FROM N.A., AND SEQUENCE OF 625-641.
 RX MEDLINE=90103465; PubMed=2604391;
 RA Moser B., Gilkes N.R., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;
 RT "Purification and characterization of endoglucanase C of Cellulomonas
 RT fimi, cloning of the gene, and analysis of in vivo transcripts of the
 RT gene.";
 RL Appl. Environ. Microbiol. 55:2480-2487(1989).
 RN [3]
 RP CELLULOSE-BINDING DOMAINS.
 RX MEDLINE=92269585; PubMed=1375311;
 RA Coutinho J.B., Gilkes N.R., Warren R.A.J., Kilburn D.G.,
 RA Miller R.C. Jr.;
 RT "The binding of Cellulomonas fimi endoglucanase C (Cenc) to cellulose
 RT and Sephadex is mediated by the N-terminal repeats.";
 RL Mol. Microbiol. 6:1243-1252(1992).
 RN [4]
 RP IDENTIFICATION OF IG-LIKE DOMAINS.
 RX MEDLINE=97035265; PubMed=8880921;
 RA Bateman A., Eddy S.R., Chochia C.;
 RT "Members of the immunoglobulin superfamily in bacteria.";
 RL Protein Sci. 5:1939-1942(1996).
 RN [5]
 RP STRUCTURE BY NMR OF 33-184.
 RX MEDLINE=97074498; PubMed=8916925;
 RA Johnson P.E., Joshi M.D., Tomme P., Kilburn D.G., McIntosh L.P.;
 RT "Structure of the N-terminal cellulose-binding domain of Cellulomonas
 RT fimi Cenc determined by nuclear magnetic resonance spectroscopy.";
 RL Biochemistry 35:14381-14394(1996).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENOGLUCANOLYDOLASES THAT CUT THE CELLULOSE POLYMER CHAIN;
 CC (2) EXOCELLULOBIOTRANSFERASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.


```

QY 230 LQPFMEKAI-----REDAYVNGSVGALIGENALAKKAVGETVRLPVNG----- 276
DB 628 WPMHCBESIVWYLFSGAGNEADVHGIFYGNTYLCKGEER-----DTANLFPKSLTLL 681
QY 277 -GPNLTSFHVIGEIDKVFEGGKGNHNI-----QTLTPAGGAATEKVDVPG 328
DB 682 MNDVTGTFDV--ECITTDHYTGMMOKYVNOCOROFEDFTVYLBERTYVDVAVEWD 739
QY 329 YVLVDHAFRAFNKALGILKVEGENHEIYSHKQTDVAVLPFGAPQALDPTOE----- 381
DB 740 Y-----SPSRAMK-----ELHHLQEQNVNVF-----LDKEFFIGSKY 774
QY 382 ---APPTPAPANIQEOIKKAGKATYDSNCA---ACHQPDGKGVNAPPLANSDYLNADH 434
DB 775 KKVYVQFTDSSFREQVQR--RAEDEHGLIGLPPRIHANVDKVKVFKMNAATRY----- 828
QY 435 ARAASIVANGLSKITVNGQYVNPALASDOQIANVITYT 477
DB 829 ---SIHAGVK-----TESSTVETLP-----GEVATYTT 854

RESULT 36
LACS_TRAVE STANDARD; PRT; 527 AA.
ID LACS_TRAVE STANDARD; PRT; 527 AA.
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Laccase 5 precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
DE (uraniol oxidase) (Diphenol oxidase) (Laccase IV).
GN LCC5 OR LCCIV.
OS Trametes versicolor (White-rot fungus).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Trametes.
OX NCBI_TaxID=5325;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=52J;
RX MEDLINE=97464057; PubMed=9322748;
RA Ong E., Pollock W.B., Smith M.;
RT "Cloning and sequence analysis of two laccase complementary DNAs from
RT the ligninolytic basidiomycete Trametes versicolor.";
RL Gene 196:113-119 (1997).
CC - FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC - CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoquinone + 2
CC H(2)O.
CC - COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Secreted.
CC - SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC - SIMILARITY: Contains 3 plastocyanin-like domains.
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U44431; AAC9829.1; -
CC DR HSSP: P37064; IAOZ.
CC DR InterPro: IPR001117; Cu-oxidase.
CC DR InterPro: IPR002355; MultiCu oxidase2.
CC DR Pfam: PFO0394; Cu-oxidase; 3.
CC DR PROSITE: PS00079; MULTICOPPER OXIDASE1; 1.
CC DR PROSITE: PS00080; MULTICOPPER OXIDASE2; FALSE_NEG.
CC DR Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
CC Lignin degradation; Multigene family.
CC KM LIGNIN DEGRADATION; Multigene family.
CC FT SIGNAL 1 23 POTENTIAL.

```

```

FT CHAIN 24 527 LACCASE 5.
FT DOMAIN 25 150 PLASTOCYANIN-LIKE 1.
FT FT 306 162 PLASTOCYANIN-LIKE 2.
FT DOMAIN 373 498 PLASTOCYANIN-LIKE 3.
FT METAL 87 87 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 89 89 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 132 132 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 134 134 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 425 425 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 428 428 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 430 430 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 480 480 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 481 481 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 482 482 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 486 486 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE . 527 AA; 56094 MW; D9597491F1F79825 CR664;

```

```

Query Match 4.1%; Score 108; DB 1; Length 527;
Best Local Similarity 20.2%; Pred. No. 3.4;
Matches 104; Conservative 62; Mismatches 190; Indels 158; Gaps 26;

```

```

QY 60 PYIDALVTHAREPPVRBDRHAKVYVVKMETEKRLADSVYQVFTGCGVPGQMIV 119
DB 27 PVTDLTISNADVTP-----DGITRAVLAGVFPFGLITG 61
QY 120 RSGDTEVO-FSNHPSKKPHNVDPH-----AATGGGAAASFT-----APGHSTPS 167
DB 62 NKGFDEQIVNVLNLTNETYKSTTHHGIPOAGTNMADG--AAFNOCPIATGNSFLYD 119
QY 168 FRAL-OPGLVYVHCAVAPVGMHIANMGV-LTLVEPK---GLPKVDEYVYMO--GDFY- 220
DB 120 FTVPPQAGTFWYH-----SHLSTQYCGDLGRLVYVYVDDNANSLYVDVDDTITLADWYH 176
QY 221 TKGKYGEOGLQPFMEKAIREDAYVYV-----FNGSVGALIGENALAKKAVE--TVRLF 272
DB 177 TAAKLG-----PAFPAGPDSVLINGLGRFSGDGGATNLTVITVQKRYAFRLV 226
QY 273 VENGPNLTSFHVIGEIFDKVFEGGKGNHNIQTLTPAGGAATEKVDVPG----- 327
DB 227 SISCPNFTFTS-----DGHNM--TILEVG--VNHKALVDSIQIRA 265
QY 328 ---DYVL-----VDHAFRAF-NKGAL-----GLKVEGENHEIYSHKQTDVAVL 369
DB 266 GGRYSFTLANGSIDINWYIRALPNGTDTTGCVNSAIIKRYTAEIEPTNATTSVILP 325
QY 370 PGCAPOALDTQCAKTP-----APANLQEOIKYG 398
DB 326 TETDLVPLDNPAPAPDPOVGYDLMSLDFSFGNSFPINNETFVPPVPL--QILSG 383
QY 399 ---KATVNSNCAACHQPDGKGVNAPPLANSDYLNADHARAASIVANGLSKITVNGNQ 455
DB 384 AODASLLEPNGSVYTLPSNSTIEISFPLTTDGLANAP-----GAPHPHLHGHT 433
QY 456 YESVNPALASDOQIANVITYTLNFGNKGGLS 489
DB 434 F-SVSRASGSSFTFNANPRADTVTSGNSGDVNT 466

RESULT 37
LACC2_AGABI

```

ID LAC2 AGABI STANDARD; PRT; 520 AA.
 AC 012542;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Laccase II precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
 DE (urithiol oxidase) (Diphenol oxidase).
 GN LCC2.
 OS Agaricus bisporus (Common mushroom).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Agaricaceae; Agaricus.
 ON NCBI_TaxID=5341;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=D649;
 RX MEDLINE=93367392; PubMed=8360614;
 RA Ferry C.R., Smith M., Britnell C.H., Wood D.A., Thurston C.F.;
 RT Identification of two laccase genes in the cultivated mushroom
 RT Agaricus bisporus.";
 RL J. Gen. Microbiol. 139:1209-1218(1993).
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoemiquinone + 2
 CC H(2O).
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L10663; AAA17035.1; -.
 DR HSSP; P37064; IASQ.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; MultiCu oxidase2.
 DR Pfam; PF00394; Cu-oxidase_3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 DR Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 KW Glycoprotein; Repeat; Multigene family.
 FT SIGNAL 1 19
 FT CHAIN 20 520 LACCASE II.
 FT DOMAIN 21 145 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 157 305 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 375 488 PLASTOCYANIN-LIKE 3.
 FT METAL 82 82 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 84 84 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 127 127 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 129 129 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 417 417 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 420 420 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 422 422 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 470 470 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 471 471 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 472 472 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 476 476 COPPER (TYPE 1) (BY SIMILARITY).
 FT CARBOHYD 108 108 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 241 241 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 454 454 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 492 492 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 520 AA; 57822 MW; EF100145F0A697AA CRC64;

Query Match 4.1%; Score 107.5; DB 1; Length 520;
 Best Local Similarity 20.5%; Pred. No. 3.6;

Matches 78; Conservative 43; Mismatches 120; Indels 139; Gaps 21;

QY 96 RLA-DGVEYFMTFGGVGCMIRVEGDTIEVQSPSNH-PDSKMHNVDFH-----AA 146
 DB 32 RLAPGDFEEDTVYVNGEFGTIVVNGKSVRI PVNNKLTSTMRKRSYIHNGHGFQART 91
 QY 147 TCGPGGAESAF--TAPGHTSFSPK-ALQGLVYHCAVAPVGMHIANGVGLIV---- 199
 DB 92 SQGDGPAPVNGCPQPPNTFTYEFVSADSGTFWYH---SHLSTGYCDGLRGAFFVYDPE 148
 QY 200 EKEGELPKYDK-----YYVMQDPYTKGKGEQGLQFPDMEKAIREDAEVVFNG 250
 DB 149 DPLGLHYVDDETTVITLAEMWYHVLAPDI--NNEFFSSGILIV-----QDS----- 192
 QY 251 SVGALTGENALAKAGETVRLFVNGGPNLTSSFVHIGEIFDKVPEGGK----- 301
 DB 193 --GLNGKGRF-----NGGPEPTP-----FAVNVEQGRKRRFVIALS 228
 QY 302 -----ENHNIQTLIPAGGAAITERKVDVPGDYVLVDHAFRAFNKALGILKVEG 352
 DB 229 CRPFTFSVDNNL--TFMEAD-----SVEHDPVEI----- 257
 QY 353 EENHEIYSHKQTDVAYLPEGAPQALDT--QEAPEKT---PAPANIQEQIKAGKATYSNC 406
 DB 258 -GNVDIYAAQVSVIL--NANQPDVNWWRAPMTGNGPNRPNINISLTIALRY---- 309
 QY 407 AACHPDGGKGVNAPFPLAN 426
 DB 310 -----NGAPEVEPTTVN 321

RESULT 38
 ID CCAA_BACTU STANDARD; PRT; 1257 AA.
 AC 045754;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pesticidal crystal protein cry12Aa (insecticidal delta-endotoxin
 DE CryIIAa) (Crystalline entomocidal protoxin) (142 kDa crystal
 DE protein)
 GN CRY12AA OR CRYIIA(A) OR CRYVB.
 OS Bacillus thuringiensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1428;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.
 RA STRAIN=NRRL B-18244 / PS33F2;
 RC Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,
 RA Sick A.J.;
 RT "Novel Bacillus thuringiensis microbes active against nematodes, and
 RT genes encoding novel nematode-active toxins cloned from Bacillus
 RT thuringiensis".
 RL Patent number EP0462721, 27-DEC-1991.
 CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.
 CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
 CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
 CC OF THE SPORE COAT.
 CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
 CC N-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L07027; AAA22355.1; -.
 DR InterPro; IPR005638; endotoxin_C.
 DR InterPro; IPR005639; endotoxin_N.

DR Pfam: PF03944; endotoxin C; 1.
 DR Pfam: PF03945; endotoxin N; 1.
 KW Toxin; Sporulation.
 SQ SEQUENCE 1257 AA; 142265 MW; 3D988F6C0E3961 CRC64;

Query Match 4.1%; Score 107.5; DB 1; Length 1257;
 Best Local Similarity 19.2%; Pred. No. 11;
 Matches 118; Conservative 69; Mismatches 200; Indels 227; Gaps 32;

25 NOADKAQKPSSTVDAAKTAANADNAAQSEHOGELVDAIATVHADEPPVDRDPAPV 84
 245 NGIDRF---KSLDVNSYKNAKVIKMT-----MVLDA-----LPTEDPDHYQKE 290
 QY 85 VVKMETVEKRLADGVEYQ-----FWTEGQVPCQMIRVEGDITIEVQSNHPDS--- 135
 DB 291 V---EIEFRTTISPT-YQVPKKNQNTSSSIVPSDLFHY-QGDLVKLEFSTRDNDGL 344
 QY 136 -----KMP-----HNVDFHAAIGPBGGAASFTAP----- 160
 DB 345 AKIFGTGIRTFYKSPNTHETHYVDSYNTQSSGNISRGSSNPIDLNPIISTCIRNSF 404
 QY 161 -----GHTSFSPFKALQPLGYVYHCAPVG-----MHANGMY----- 194
 DB 405 YKAIAGSSVUNFK--DGTQYAFAPQAPFTGAMHDSFIESDGAPEGHKLTNTIYISPGDT 461
 QY 195 -----GLILVEPKEGLPKVDKEYVMQG--DFTYKGYGEQGLQPFDM 236
 DB 462 LADFNVYTLSTPTINELSTEXIKGF-AEKGYINQGIKMYGKPE-INGAPVNI 519
 QY 237 KA-----IREDAEYV-----FNGSVGALTGNA-----LKAKVET 268
 DB 520 NOOTLIFEFHAKTKQYTRIRIRYASTQYGRFLDNLQTLNIPSHNGYVATGNGEN 579
 QY 269 VLFVFGNGGPNLTSSPHVI-----GEIFPKVHEGSGKGNHNIQTLLPAGGAITE 320
 DB 580 YDLVY-IGSYTTEGNTTQIQHNDKGNVLDRIEF-----VPSDLO-- 621
 QY 321 FKVDVPGDVVLVDHAIFFRAFNKALGILKEGSENEIYSHKQTDVAVLPEGAPOAIDTQ 380
 DB 622 ---DSFQDSPPREVHSTTIFDKSSPTI---WSSNHSYSHLHBSYSYSSQS----- 667
 QY 381 EAPKTPAPANLOEQIKAGKATYDSCACHPDGKGVNAPFPL-ANSDYLNADHAPRAS 439
 DB 668 -----YFHNLLINL-----FHPID-----PARNHTIHVNNDMNVYDGKDS- 703
 QY 440 IVANGL-----SGKIT-----VNGQVESMPALASDOIAIVITYTLNS 480
 DB 704 -VADGLNENKTATIPSDAWYSGTITSMHLEFNDNNFKITTPKEELSN-ELENITTOVNAL 761
 QY 481 FGNKGQSADDAVA 494
 DB 762 FASSAQDTLASVVS 775
 RESULT 39
 FLNA_HUMAN STANDARD; PRT; 2647 AA.
 AC P21333;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Filamin A (Alpha-filamin) (filamin 1) (Endothelial actin-binding protein) (ABP-280) (Nonmuscle filamin).
 DE FLNA OR FLN1 OR FLN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eultheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=90361737; PubMed=2331361;
 RA Gorlin J.B., Yamah R., Egan S., Stewart M., Stossel T.P.,
 KWiatkowski D.J., Hartwig J.H.;

RT "Human endothelial actin-binding protein (ABP-280, nonmuscle
 RT filamin): a molecular leaf spring.";
 RL J. Cell Biol. 111:1089-1105(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96311563; PubMed=8733135;
 RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,
 RA Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlesinger D.,
 RA D'Urso M.;
 RT "Long-range sequence analysis in Xq28: thirteen known and six
 RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
 RT G6PD loci.";
 RL Hum. Mol. Genet. 5:659-668(1996).
 RN [3]
 RP SEQUENCE OF 1658-1772 FROM N.A.
 RX MEDLINE=93357748; PubMed=7689010;
 RA Maestrini E., Patroaso C., Mancini M., Rivella S., Rocchi M.,
 RA Ripetto M., Villa A., Fratini A., Zoppe M., Vezzoni P.,
 RA Toniolo D.;
 RT "Mapping of two genes encoding isoforms of the actin binding protein
 RT ABP-280, a dyatrophin like protein, to Xq28 and to chromosome 7.";
 RL Hum. Mol. Genet. 2:761-766(1993).
 RN [4]
 RP REVIEW.
 RX MEDLINE=21234905; PubMed=11336782;
 RA van der Flier A., Sonnenberg A.;
 RT "Structural and functional aspects of filamins.";
 RL Biochim. Biophys. Acta 1538:95-117(2001).
 RN [5]
 RP REVIEW.
 RX MEDLINE=21146932; PubMed=11252955;
 RA Stossel T.P., Condeelis J., Cooley L., Hartwig J.H., Noegel A.,
 RA Schleicher M., Shapiro S.S.;
 RT "Filamins as integrators of cell mechanics and signalling.";
 RL Nat. Rev. Mol. Cell Biol. 2:138-145(2001).
 RN [6]
 RP VARIANTS PH PHE-656 AND THR-1764.
 RX MEDLINE=21423782; PubMed=11532987;
 RA Sheen V.L., Dixon P.H., Fox J.W., Hong S.E., Kinton L., Sisodiya S.M.,
 RA Duncan J.S., Dubau F., Scheffer I.E., Schachter S.C., Walner A.,
 RA Henry R., Crino P., Kanuro K., Dimario F., Berg M., Kuznietchy R.,
 RA Cole A.J., Bromfield E., Biber M., Schomer D., Whales J., Silver R.,
 RA Mochida G.H., Berkovic S.F., Andermann F., Andermann E., Dobyns W.B.,
 RA Wood N.W., Walsh C.A.;
 RT "Mutations in the X-linked filamin 1 gene cause periventricular
 RT nodular heterotopia in males as well as in females.";
 RL Hum. Mol. Genet. 10:1775-1783(2001).
 CC [1]
 CC FUNCTION: Promotes orthogonal branching of actin filaments and
 CC links actin filaments to membrane glycoproteins. Anchors various
 CC transmembrane proteins to the actin cytoskeleton and serves as a
 CC scaffold for a wide range of cytoplasmic signaling proteins.
 CC [2]
 CC SUBUNIT: Homodimer. Interacts with cVsp. Interacts with various
 CC other binding partners in addition to filamentous actin.
 CC [3]
 CC TISSUE SPECIFICITY: Ubiquitous.
 CC [4]
 CC PTM: PHOSPHORYLATION EXTENT CHANGES IN RESPONSE TO CELL
 CC ACTIVATION.
 CC [5]
 CC PTM: The N-terminus is blocked.
 CC [6]
 CC DISEASE: Defects in FLNA are the cause of periventricular
 CC heterotopia (PH) also called nodular heterotopia, bilateral
 CC periventricular (NHP or BPNH). PH is an X-linked developmental
 CC dominant disorder in which many neurons fail to migrate into the
 CC cerebral cortex. They remain as nodules lining the ventricular
 CC surface. In heterozygous females these neurons presumably
 CC represent those cells that, after X-chromosome inactivation,
 CC contain the active X chromosome with the filamin mutation. Most
 CC heterozygous affected males die early during embryogenesis, whereas
 CC heterozygous females have normal intelligence but suffer from
 CC seizures and various manifestations outside the central nervous
 CC system, especially related to the vascular system. This implies
 CC that essential embryonic cell migration can only occur in FLNA-
 CC expressing cells.
 CC [7]
 CC SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY

CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
 CC ABP-120, ABP-180, OR BETA-FODRIN).
 CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -1- SIMILARITY: Contains 24 filament repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X53416; CAA37495.1; -
 DR EMBL: U44140; AAA92644.1; -
 DR EMBL: X70082; CAA96687.1; -
 DR EMBL: X70085; CAA96690.1; -
 DR PIR: A37098; A37098.
 DR HSSP: P13466; 1KSR.
 DR Gene; HGNC:3754; FLNA.
 DR MIM: 300017; -
 DR MIM: 300049; -
 DR GO: GO:0015629; Cactin cytoskeleton; TAS.
 DR GO: GO:0003780; F-actin cross-linking activity; TAS.
 DR GO: GO:0007012; F-actin cytoskeleton reorganization; TAS.
 DR GO: GO:0007399; F-actinogenesis; TAS.
 DR InterPro: IPR001589; Actbind_actin.
 DR InterPro: IPR001298; Calponin-like.
 DR Pfam: PF00307; CH; 2.
 DR Pfam: PF00630; Filamin; 24.
 DR SMART: SM00033; CH; 2.
 DR SMART: SM00557; IG_FLMN; 24.
 DR PROSITE: PS00019; ACTININ_1; 1.
 DR PROSITE: PS00020; ACTININ_2; 1.
 DR PROSITE: PS50021; CH; 2.
 DR PROSITE: PS50194; FILAMIN_REPEAT; 24.
 DR Actin-binding; Phosphorylation; Repeat; Polymorphism;
 DR disease mutation.
 KW
 FT DOMAIN 1 274 ACTIN-BINDING (HEAD).
 FT 43 149 CH 1.
 FT DOMAIN 266 374 CH 2.
 FT REPEAT 276 374 FILAMIN 1.
 FT REPEAT 376 474 FILAMIN 2.
 FT REPEAT 475 570 FILAMIN 3.
 FT REPEAT 571 663 FILAMIN 4.
 FT REPEAT 667 763 FILAMIN 5.
 FT REPEAT 764 866 FILAMIN 6.
 FT REPEAT 867 965 FILAMIN 7.
 FT REPEAT 966 1061 FILAMIN 8.
 FT REPEAT 1062 1154 FILAMIN 9.
 FT REPEAT 1155 1249 FILAMIN 10.
 FT REPEAT 1250 1349 FILAMIN 11.
 FT REPEAT 1350 1442 FILAMIN 12.
 FT REPEAT 1443 1539 FILAMIN 13.
 FT REPEAT 1540 1636 FILAMIN 14.
 FT REPEAT 1649 1740 FILAMIN 15.
 FT REPEAT 1741 1778 HINGE 1.
 FT DOMAIN 1779 1860 FILAMIN 16.
 FT REPEAT 1861 1950 FILAMIN 17.
 FT REPEAT 1951 2039 FILAMIN 18.
 FT REPEAT 2042 2131 FILAMIN 19.
 FT REPEAT 2132 2230 FILAMIN 20.
 FT REPEAT 2233 2325 FILAMIN 21.
 FT REPEAT 2327 2420 FILAMIN 22.
 FT REPEAT 2424 2516 FILAMIN 23.
 FT DOMAIN 2517 2551 HINGE 2.
 FT REPEAT 2552 2646 FILAMIN 24.
 FT DOMAIN 2647 2761 SELF-ASSOCIATION SITE, TAIL.
 FT SITE 1761 1762 CLEAVAGE (BY CALPAIN).
 FT VARIANT 320 370 V -> A (IN dbSNP:1064816).
 FT VARIANT 370 370 F -> L (IN dbSNP:1064817).

FT FT
 FT VARIANT 552 552 /FTid=VAR 012832.
 FT VARIANT 656 656 V -> A (IN dbSNP:730319).
 FT VARIANT 656 656 /FTid=VAR 012833.
 FT VARIANT 1764 1764 L -> F (IN PH).
 FT VARIANT 1764 1764 /FTid=VAR 012834.
 FT CONFLICT 1772 1772 A -> T (IN PH).
 FT CONFLICT 2634 2634 /FTid=VAR 012835.
 FT CONFLICT 2634 2634 A -> G (IN REF. 3).
 SQ SEQUENCE 2647 AA; 280759 MW; 6C1A07041DPA3D42 CRC64;
 Query Match 4.1%; Score 107.5; DB 1; Length 2647;
 Best Local Similarity 20.8%; Pred. No. 30;
 Matches 115; Conservative 71; Mismatches 217; Indels 149; Gaps 30;
 QY 4 PTLITITICALSALMLSGCNOADKAQPSSTYDAAKTANMDNASSOHOSELPYID 63
 DB 2151 PSVANVGSCHDLS-LKPEISIQ-DMTAQVTS--PSGKTHEIVEGENTTYCIRFPV 2204
 QY 64 AIV-----THAPEVP-----PVDRDHPAKVVKMETVEKVRADGV--EYOF 105
 DB 2205 AEMGHTIVSVKKGQHVGSFPQFTVGPLGEGAKVAGRGLE--RAEGVPAEFSI 2261
 QY 106 WTFGGQVPGQMRVREGDTIEVQFSNHPDS-----KMPHNVFHAATPGGGAENSF 157
 DB 2262 WTRBAGAGGLAIVGSPKAEISFEDRKDSCGVAVVQVEPDYEVSVKFNENHIDSPF 2321
 QY 158 TAP-----GHSTSEFKALQ-PGLVYVYCAVAPVGMHIANGVGILVEPKGPKVDKE 211
 DB 2322 VFPVAPSGDARRLVSSLSQSGKLVNQPAFPAVSLNKAIGALIDAKVHSPGAL---EE 2377
 QY 212 YVWQGDPTKKGKVGEGGLQPFDEKAIREDAEVYV---FNGSVGLTGENALKAKVGET 268
 DB 2378 CYVTED---QDKVAVRPIP-----RENGVILVDKFNCT-HIG-SPFKRVGEP 2423
 QY 269 VRLFVNGG-PULTSSFVIGEIFDKVREGGKGN--HNIOITLIPAGCAIT----- 319
 DB 2424 -----GHGGDPLVSAV-----GAGLEGVTNPAPFVNTSNAGAGALSVTIDGPS 2470
 QY 320 EKKVD-----VPGDYV-----VDHATFRANKKALGILKKEGE-- 353
 DB 2471 KYKMDCOBCEPREGVRYTPMAPGSYLISIKYGGPHIGSPKKA-----KVTGPR 2521
 QY 354 -ENHEIYSHKOTDAVYLPEGAPOALDIOEAPKTPAPMLQEOIKAG-----KATY 402
 DB 2522 VSNHSLH--ETSSVFSVSLTATCAPOHGAAGRPADASKVYVAGLGSKAVYQKSSF 2578
 QY 403 DSNCA-----ACHOP-----DGKGVNAPFPPLANSDYLNADHARAASIYANGLSGI 449
 DB 2579 TVDCSKAGNNMLVGVHGRTCEILVGHVGSRLYSVYLKDGGEYTLVVK--WGHE 2635
 QY 450 TVNGNOYESVMP 461
 DB 2636 HIRGSPYRVVP 2647
 RESULT 40
 OMPB_RICRI STANDARD; PRT; 1654 AA.
 AC 053047;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (tompB)
 DE (tomp B) (Contains: 120 kDa surface-exposed protein (surface protein
 DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
 GN OMPB
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxId=783;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=92267802; PubMed=1724278;
 RA Gilmore R.D. Jr., Cieplik W. Jr., Pollicastro P.F., Hackstadt T.;
 RT "The 120 kD surface-exposed protein of Rickettsia rickettsii
 RT rickettsii is encoded by an unusually long open reading frame:
 RT evidence for protein processing from a large precursor.";
 RL Mol. Microbiol. 5:2361-2370(1991).
 RP SEQUENCE OF 279-1654 FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=90136087; PubMed=2515418;
 RA Gilmore R.D. Jr., Joete N., McDonald G.A.;
 RT "Cloning, expression and sequence analysis of the gene encoding the
 RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
 RL Mol. Microbiol. 3:1579-1586(1989).
 CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X16353; CAA34403.1; -
 CC PIR: S18227; S18227.
 DR InterPro: IPR006315; Autotransport.
 DR InterPro: IPR005546; Autotransporter.
 DR Pfam: PF03797; Autotransporter; 1.
 DR TIGRPFAMs: TIGR01414; autotrans_bar1; 2.
 DR Antigen: S-layer; Cell wall.
 FT CHAIN 1 1333 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1334 1654 32 kDa BETA PEPTIDE.
 FT DOMAIN 1181 1188 POLY-THR.
 SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 4.1%; Score 107; DB 1; Length 1654;
 Best Local Similarity 20.9%; Pred. No. 18;
 Matches 128; Conservative 65; Mismatches 228; Indels 190; Gaps 29;

QY 6 LIKTLILCALSLMSSGCSNOADKAAQPKSSTVDAAKT--ANADNAASQEHQGLPVI 62
 DB 11 LISAGLVSTATVIVASFAGSAMGAIQONRTTNGAATVVDGAFDQTAAPAVG--VAL 68
 QY 63 DAIVTHAPEVPPVDRDHPA-----KVVVKMEYVEKMRLLADGVEYQFTF 108
 DB 69 NNAVIT--ANANGINENTPAGSPNGLLTANNLAATVASEDT--TLGFTTVVNHANSP 123
 QY 109 G-----GOVPGQMI RVEGDTIEVOFSN-----HPDSKMPHNVDFHAATGP 149
 DB 124 NLTLAGKTLITTGCVTAQAQAATKNAQNVVQFNNGAIDNNDLKVGRIIDF----- 177
 QY 150 GGGAEASFTAPGHTSTFSFKALQPGLYVYHCAVAPV---GWHIANGMYGLILVEPKGL 205
 DB 178 --GAPV-----STLVFNLANP-----TQKAPLILGDNAVIANGVNGTLNV----- 216
 QY 206 PAVDEKYYVGGDFYTKGKYGGGGLQPFDEKAIR--EDAEVYVNGSVGALTGENALAKK 264
 DB 217 -----TNG--FIOVSNKSFATVAKAINIADGGIIPNTDA--NNANTLNQ 257
 QY 265 VGEIVTLFVNGGPN-----LTSFPHVIGEI-----PDKVHFEKGKGENHNI 306
 DB 258 AGTTINFTGTGTGTLVLSKGAATNTNITGSLGNLKVIEFTVAVDQGLTANAGA 317
 QY 307 QTTLLI---PAGGAATFERKVDVPGDYVLVDHAIFFAENKGAIGLIKVEGE---ENHEIY 359

DB 318 ANAVIGTNNAGRAA-----GFVVSVDNG-----KVATIDGOVYAKDMVIO 358
 QY 360 SHKQTDVAVYLPFGAPQALDPTQAPKTPAP-----ANLQEQIK----- 396
 DB 359 SANATGOVNFRIIVDVGADGTTAFRTAASKVITTDQSNFGNDFGLAAGIKVPNAITLT 418
 QY 397 ---AGKATYDSNCAACHQDQKGV---PNAFPPLANSDYLANDHAPAAISV----- 441
 DB 419 GNFTGDASNPGNTAGVITFDANGTLESASADANVAVTNNITALEASGAGVQLSGTHAAE 478
 QY 442 -ANGUSGKI-----TVNG--NOYESVVPALALSDQOLANVITTTLNSFGKKGGLSLAD 491
 DB 479 LRLGNAGSIFKLADGTVINGKVNQOTALVGALAAAGTTLDGSAITTT-GDIGNAGGAAALQ 537
 QY 492 DVA---KAKKT 499
 DB 538 RITLANDAKKT 548

Search completed: August 27, 2003, 18:30:06
 Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:33:27 ; Search time 26 Seconds

(without alignments)
907.977 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502
Sequence: 1 MSKPTLIKTTLLICALSALML.....NKGQLSADDAVAKKTKPEN 502

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	6.8	392	1 ANIA_NERGO	Q02219 neisseria g
2	12	2.4	330	1 NIR_ALCKX	P81445 alcaligenes
3	12	2.4	363	1 NIR_PSECL	Q06006 pseudomonas
4	10	2.0	377	1 NIR_RHINE	O60214 rhizobium h
5	8	1.6	134	1 RCSF_ECOLI	P28633 escherichia
6	8	1.6	389	1 PORA_MERJA	O57715 methanococ
7	8	1.6	432	1 TIG_THERN	O8r26 chernomae
8	8	1.6	518	1 SAP_CHICK	O13035 gallus gall
9	8	1.6	604	1 YFIC_BACSU	P54719 bacillus su
10	7	1.4	86	1 C555_CHILT	P00123 chlorobium
11	7	1.4	108	1 C555_CHLTE	O8K693 chlorobium
12	7	1.4	109	1 RBS_PROHO	P27569 prochloroth
13	7	1.4	111	1 RBS_SYN2	O44178 synechococ
14	7	1.4	111	1 YC39_AQUAE	O67284 aquifex aeo
15	7	1.4	130	1 RS9_LACLA	O9ydg7 lactococcus
16	7	1.4	157	1 GREB_BRUME	O8yid6 bruceella me
17	7	1.4	172	1 KCH5_MOUSE	O92063 mus musculu
18	7	1.4	187	1 BEA3_MOUSE	O08545 mus musculu
19	7	1.4	193	1 YEAY_ECOLI	P76235 escherichia
20	7	1.4	194	1 R1SE_ARCFU	O27965 archaeoglob
21	7	1.4	212	1 RPSD_MYCTU	O50712 mycobacteri
22	7	1.4	218	1 YIDX_ECOLI	P1461 escherichia
23	7	1.4	226	1 TRIS_MERTR	O74025 methanobact
24	7	1.4	230	1 YX09_CAEEL	O11115 caenorhabdi
25	7	1.4	238	1 EFA3_HUMAN	P52737 homo sapien
26	7	1.4	239	1 YGIP_YEAST	P53223 saccharomye
27	7	1.4	272	1 FWDG_METMP	O31112 methanococ
28	7	1.4	319	1 SYGA_COXBU	P94616 coxiella bu
29	7	1.4	328	1 HAM1_STRP3	O8K847 streptococc
30	7	1.4	328	1 HAM1_STRP8	O8P242 streptococc
31	7	1.4	328	1 HAM1_STRPY	O8A186 streptococc
32	7	1.4	351	1 PYRB_LACLE	O60252 lactobacill
33	7	1.4	352	1 AVRC_PSESG	P13836 pseudomonas

34	7	1.4	374	1 METL_PEA	P49613 pisum sativ
35	7	1.4	374	1 NIR_RHOSH	Q53239 rhodobacter
36	7	1.4	376	1 NIR_ALCEA	P38501 alcaligenes
37	7	1.4	378	1 NIR_ACHCY	P25006 achromobact
38	7	1.4	378	1 YMA2_MYCTU	O10400 mycobacteri
39	7	1.4	387	1 REC2_LACLA	O01840 lactococcus
40	7	1.4	393	1 METK_CATRO	O96551 catnaranthu
41	7	1.4	393	1 METK_PINBN	P50300 pinus banks
42	7	1.4	393	1 METL_LYCES	P43281 lycopersico
43	7	1.4	404	1 KAPR_COLTR	O42794 colletoeric
44	7	1.4	436	1 GFP6_BOVIN	P55106 bos taurus
45	7	1.4	444	1 TIG_FROCA	O68129 rhodobacter
46	7	1.4	459	1 YGWS_YEAST	P53083 saccharomye
47	7	1.4	462	1 LEU2_MOUSE	O08580 mus musculu
48	7	1.4	462	1 LEU2_LISIN	O92a26 listeria in
49	7	1.4	502	1 ATPA_BACHD	O9K6h3 bacillus ha
50	7	1.4	508	1 RPA2_THEVO	O979E5 thermoplas
51	7	1.4	512	1 FUS1_YEAST	P11710 saccharomye
52	7	1.4	519	1 ERRI_HUMAN	P11474 homo sapien
53	7	1.4	525	1 AROF_ARATH	P29976 arabidopsis
54	7	1.4	550	1 YAO8_SCHPO	O10087 schizosacch
55	7	1.4	571	1 SYE_METKA	O8rxb7 methanopyru
56	7	1.4	586	1 YAGM_RHISN	P5469 rhizobium s
57	7	1.4	595	1 SYD_AGRIS	O8u87 agrobacteri
58	7	1.4	610	1 GLWS_PSEAE	O9h25 p glucosami
59	7	1.4	615	1 RPSD_PSEFL	P52326 pseudomonas
60	7	1.4	633	1 BZ21_YEAST	P38822 saccharomye
61	7	1.4	664	1 PH2M_TRICU	P15245 trichosporo
62	7	1.4	706	1 Z151_CHICK	O90625 gallus gall
63	7	1.4	733	1 MUTB_STRCM	O05065 streptomyc
64	7	1.4	807	1 MCW3_XENLA	P49739 xenopus lae
65	7	1.4	859	1 ALR1_YEAST	O08269 saccharomye
66	7	1.4	962	1 KCH1_RAT	O63472 rattus norv
67	7	1.4	987	1 KCH1_BOVIN	O18965 bos taurus
68	7	1.4	988	1 KCH5_HUMAN	O8ncm2 homo sapien
69	7	1.4	988	1 KCH5_RAT	O9ep19 rattus norv
70	7	1.4	989	1 KCH1_HUMAN	O95259 homo sapien
71	7	1.4	989	1 KCH1_MOUSE	O60603 mus musculu
72	7	1.4	995	1 YPDI_CAEEL	P48053 caenorhabdi
73	7	1.4	1030	1 PEK6_YEAST	P33760 saccharomye
74	7	1.4	1174	1 CIKE_DROME	O02280 drosophila
75	7	1.4	1174	1 KCRP_STRPU	P18280 strongyloce
76	7	1.4	1331	1 Y064_MYCGE	P47310 mycoplasma
77	7	1.4	1596	1 MAM_DROME	P21519 drosophila
78	7	1.4	3011	1 YOS1_SCHPO	P81719 schizosacch
79	7	1.4	3080	1 POLG_ZMYVC	P18479 z genome po
80	7	1.4	3083	1 POLG_ZMYVR	O89330 z genome po
81	7	1.4	3083	1 POLG_ZMYVS	O36979 z genome po
82	7	1.2	33	1 ANP3_MYOAC	P04367 myoxocephal
83	7	1.2	33	1 ANP5_MYOAE	P20421 myoxocephal
84	7	1.2	33	1 AKH2_SCHGR	P35808 schistoserc
85	7	1.2	66	1 RS27_SULTO	O92749 sulfolobus
86	7	1.2	72	1 FMT_RICE	O33523 rickettsia
87	7	1.2	72	1 Y1IF_ECOLI	P32150 escherichia
88	7	1.2	73	1 FMT_RICAM	O33508 rickettsia
89	7	1.2	73	1 FMT_RICPA	O33543 rickettsia
90	7	1.2	73	1 FMT_RICRH	O33545 rickettsia
91	7	1.2	73	1 FMT_RICRI	O33544 rickettsia
92	7	1.2	73	1 FMT_RICSI	O33575 rickettsia
93	7	1.2	73	1 KIL_ECOLI	P38393 escherichia
94	7	1.2	83	1 CYC6_PAVLU	P00107 pavlova lut
95	7	1.2	85	1 CYC6_PLEBO	P00117 plectonema
96	7	1.2	85	1 CYC6_PORTE	P00111 porphyra te
97	7	1.2	86	1 CYC6_BUMFI	P00110 bunilleriop
98	7	1.2	87	1 CYC6_STNLI	P00114 synechococ
99	7	1.2	87	1 CYC6_STNLI	P00115 synechococ
100	7	1.2	88	1 CYC6_BRYMA	P11448 bryopsis ma
101	7	1.2	88	1 Y073_BACHD	O9K518 bacillus ha
102	7	1.2	89	1 CYC6_MONBR	O07969 monoraphidi
103	7	1.2	89	1 CYC6_SCEOB	P57739 scenedemus
104	7	1.2	89	1 CYC6_SPIMA	P00118 spirulina m
105	7	1.2	91	1 CYC6_CIAGO	P83391 cladophora
106	7	1.2	94	1 N18W_NEUCR	O07842 neurospora

107	6	1.2	95	1	RR20_CVAPA	P48140	Cyanophora	180	6	1.2	149	1	FUR_VIBPA	024755	vibrio para
108	6	1.2	97	1	CH10_BUCCU	Q9444	buchnera ap	181	6	1.2	149	1	FUR_VIBPA	P33117	vibrio vuln
109	6	1.2	97	1	GATC_LISIN	Q92a22	listeria in	182	6	1.2	150	1	FUR_VIBCH	P33087	vibrio chol
110	6	1.2	97	1	GATC_LISMO	P58817	listeria mo	183	6	1.2	150	1	VC18_VACC	P23102	vaccinia vi
111	6	1.2	98	1	FER3_ANAVA	P46050	anabaena va	184	6	1.2	150	1	YM22_VIBVU	08d407	vibrio vuln
112	6	1.2	104	1	RL27_MYGE	P47476	mycoplasma	185	6	1.2	150	1	YPSL_STYEN	P25508	synecococ
113	6	1.2	105	1	NAL2_CANAL	P46594	candida alb	186	6	1.2	151	1	YX1F_BACSU	Q93298	escherichia
114	6	1.2	108	1	NIRD_ECOLI	P3675	escherichia	187	6	1.2	156	1	YX1F_BACSU	P42298	brucella su
115	6	1.2	109	1	NIRD_SALTU	P40769	salmonella	188	6	1.2	158	1	RS9_BRUME	08y910	brucella me
116	6	1.2	109	1	CYC6_CVACA	Q9141	cyamidium c	189	6	1.2	161	1	HBPL_PARAD	04396	parasponia
117	6	1.2	109	1	RLA1_MAIZE	P52855	zea mays (m	190	6	1.2	162	1	Y875_MYCTU	010537	mycobacteri
118	6	1.2	109	1	ZM33_MAIZE	082106	zea mays (m	191	6	1.2	162	1	Y9H4_PSEAE	Q9h10	pseudomonas
119	6	1.2	110	1	CYC6_PORPU	P51200	porphyra pu	192	6	1.2	163	1	C550_SYNEL	P56150	synecococ
120	6	1.2	110	1	CYC6_PORPE	Q8wK18	porphyra ye	193	6	1.2	163	1	H1_TETTH	P10156	tetrahymena
121	6	1.2	111	1	CYC6_SYNP7	P25935	synecococ	194	6	1.2	164	1	GREA_MYCLE	P4608	mycobacteri
122	6	1.2	112	1	CYC6_SYNEL	P56534	synecococ	195	6	1.2	164	1	GREA_MYCTU	053428	mycobacteri
123	6	1.2	112	1	CYC6_SYNVU	Q9f119	synecococ	196	6	1.2	164	1	YOH4_ECOL6	08f416	escherichia
124	6	1.2	112	1	RLA1_DROME	P08570	drosofila	197	6	1.2	164	1	YOH4_ECOL1	P52082	escherichia
125	6	1.2	112	1	RLA2_BABBO	P27055	babeia bov	198	6	1.2	164	1	YOH4_SALTI	Q823q8	salmonella
126	6	1.2	113	1	YG35_BPMV4	Q04775	lactococcus	199	6	1.2	164	1	YOH4_SALTI	Q823q8	salmonella
127	6	1.2	114	1	MAUL_METEX	Q49129	methylobact	200	6	1.2	166	1	PLAS_FR1AG	022646	frutillaria
128	6	1.2	114	1	RLA1_CHICK	P18660	gallus gall	201	6	1.2	166	1	RNH2_LACTC	030415	lactococcus
129	6	1.2	114	1	RLA1_MOUSE	P47955	mus musculus	202	6	1.2	167	1	YDER_ECOLI	P77294	escherichia
130	6	1.2	115	1	IF1A_PYRFU	Q8u0K5	pyrococcus	203	6	1.2	169	1	HXA9_CHICK	Q98324	gallus gall
131	6	1.2	115	1	IF1A_PYRFU	Q59280	pyrococcus	204	6	1.2	174	1	*PAP3_MOUSE	009049	mus musculus
132	6	1.2	115	1	RL19_BUCAL	P57477	buchnera ap	205	6	1.2	174	1	POPB_RALSO	Q9481	raletonia s
133	6	1.2	115	1	RNPA_BUCAL	P57130	buchnera ap	206	6	1.2	175	1	AXIK_ARATH	024410	arabidopsis
134	6	1.2	117	1	CYC6_SYNP2	Q10881	synecococ	207	6	1.2	176	1	UCRI_BRAJA	P51130	bradyrhizob
135	6	1.2	117	1	RL18_HAEIN	P44356	haemophilus	208	6	1.2	180	1	CTGI_HUMAN	P78358	homo sapien
136	6	1.2	117	1	RL18_VIBCH	Q9Kp00	vibrio chol	209	6	1.2	181	1	Y573_RICPR	Q92x88	ricketsia
137	6	1.2	117	1	RL18_VIBPR	P52863	vibrio proc	210	6	1.2	182	1	ID1_BREIN	Q9Kf75	bradyrhizob
138	6	1.2	118	1	YG35_BPLH	Q04769	lactococcus	211	6	1.2	183	1	NU5M_PEA	Q95639	pisum sativ
139	6	1.2	120	1	CYC6_SYNP3	P46445	synecocyst	212	6	1.2	184	1	RRF_AOUAE	065628	aquifex aeo
140	6	1.2	120	1	MIFH_CAEEL	P18785	caenorhabdi	213	6	1.2	186	1	ATPD_RHOBL	Q05437	rhodospseudo
141	6	1.2	122	1	LITH_PIG	Q29191	sus scrofa	214	6	1.2	186	1	DHML_METFL	050425	methylobact
142	6	1.2	122	1	Y590_VIBVU	Q8d6j8	vibrio vuln	215	6	1.2	186	1	TNR5_PSEAE	P06691	pseudomonas
143	6	1.2	124	1	PA21_BOTJA	P81243	bohreria ja	216	6	1.2	186	1	TNR7_ECOLI	P06692	escherichia
144	6	1.2	124	1	RL7_RALSO	Q8xue7	raletonia s	217	6	1.2	187	1	DHML_METME	059543	methylobact
145	6	1.2	127	1	RL17_XANCP	Q93466	xanthomonas	218	6	1.2	187	1	RRF_PARZE	0811h5	paracoccus
146	6	1.2	129	1	ACP2_HORVU	P08817	hordeum vul	219	6	1.2	188	1	SLP_ECOLI	P37194	escherichia
147	6	1.2	129	1	ACP2_PSEAE	Q52761	pseudomonas	220	6	1.2	189	1	ERV1_YEAST	P27882	saccharomye
148	6	1.2	130	1	RS9_STRPN	Q978n4	streptococ	221	6	1.2	189	1	GRPE_HELPY	P55970	helicobacte
149	6	1.2	130	1	RS9_STRPY	Q99y08	streptococ	222	6	1.2	191	1	GRPE_HELPY	Q92mw3	helicobacte
150	6	1.2	131	1	HS1R_HAEIN	P44754	haemophilus	223	6	1.2	193	1	RNH2_RICCN	Q92105	ricketsia
151	6	1.2	131	1	RL32_CANAL	Q94008	candida alb	224	6	1.2	194	1	CLPP_CAMJE	P54413	campylobact
152	6	1.2	132	1	RS9_MYCEN	P57179	mycoplasma	225	6	1.2	194	1	HSB_XENLA	P22845	xenopus lae
153	6	1.2	134	1	ACP1_BRANA	P10352	brassica na	226	6	1.2	195	1	NU5M_MARRO	P34944	maricantia
154	6	1.2	134	1	ACP5_BRANA	P17650	brassica na	227	6	1.2	195	1	YLM2_CAEEL	P34376	caenorhabdi
155	6	1.2	134	1	ACP5_BRANA	P08971	brassica na	228	6	1.2	196	1	H5A_XENLA	P22844	xenopus lae
156	6	1.2	134	1	ACP BRACH	P07088	brassica ca	229	6	1.2	196	1	NODA_RHIME	P02862	rhizobium m
157	6	1.2	136	1	ACP1_CASGL	P93092	casuarina g	230	6	1.2	198	1	VCO7_ADE02	P03266	human adeno
158	6	1.2	136	1	RL28_BAT	P17702	rattus norv	231	6	1.2	199	1	UREG_HELPY	Q92m77	helicobacte
159	6	1.2	139	1	PSAD_ANASP	P58573	anabaena sp	232	6	1.2	199	1	UREG_HELPY	Q09966	helicobacte
160	6	1.2	140	1	YMB4_PYRAE	Q82ct1	pyrobaculum	233	6	1.2	200	1	COWA_BRAJA	P30963	bradyrhizob
161	6	1.2	141	1	FAB2_FUSNN	Q81690	fusobacteri	234	6	1.2	202	1	Y357_CAMJE	Q9p146	campylobact
162	6	1.2	141	1	RS12_MERTH	Q27129	methanobact	235	6	1.2	203	1	YCXW_PORPU	P51360	porphyra pu
163	6	1.2	142	1	YEDD_SALTU	006399	salmonella	236	6	1.2	204	1	KADA_AERPE	Q93462	aeropyrum p
164	6	1.2	142	1	RLN1_MYCTU	P96931	mycobacteri	237	6	1.2	204	1	UREG_BACSB	007403	bacillus sp
165	6	1.2	142	1	RNPA_XYLPA	Q9P9u0	xyella fae	238	6	1.2	204	1	UREG_STRAX	P42877	staphylococ
166	6	1.2	142	1	RS12_THEAC	Q9h1y2	thermoplas	239	6	1.2	204	1	UREG_STRSL	Q50507	streptococ
167	6	1.2	142	1	RS12_THEVO	Q97cd8	thermoplas	240	6	1.2	205	1	RUVA_AGRTS	Q8uK55	agrobacteri
168	6	1.2	142	1	YE63_METTM	050770	methanobact	241	6	1.2	205	1	UREG_ECOLI	P10387	escherichia
169	6	1.2	144	1	AMSI_ERWAM	Q46630	erwinia amy	242	6	1.2	205	1	UREG_KLEAE	P18319	klebsiella
170	6	1.2	146	1	HBE2_BOVIN	P06642	bos taurus	243	6	1.2	205	1	UREG_PROMI	Q06306	procureu mir
171	6	1.2	146	1	HBE2_CAPHI	P02105	capra hircu	244	6	1.2	206	1	RUVA_RHIEI	Q919c4	thiobium e
172	6	1.2	146	1	HBE4_BOVIN	P06643	bos taurus	245	6	1.2	206	1	UREG_SYNY3	P72955	synecocyst
173	6	1.2	147	1	RS12_METKA	Q8tX12	methanopyru	246	6	1.2	206	1	UREG_UREPA	Q56361	ureaplasma
174	6	1.2	147	1	RS12_PYRAB	Q8tX10	pyrococcus	247	6	1.2	208	1	INCB_ECOLI	P05631	escherichia
175	6	1.2	147	1	RS12_PYRHO	059229	pyrococcus	248	6	1.2	209	1	LPCA_XENLA	P06692	xenopus lae
176	6	1.2	147	1	RS12_THECC	P29161	thermococ	249	6	1.2	209	1	LPCA_CHLTE	Q8k4w3	chlorobium
177	6	1.2	148	1	CS52_THETH	P04164	thermus the	250	6	1.2	210	1	CUTI_PHYCP	P41754	phytophthor
178	6	1.2	148	1	CYC6_CHLRE	P08197	chlamydomon	251	6	1.2	210	1	VP28_CAEEL	Q9a226	caenorhabdi
179	6	1.2	149	1	FUR_VIBAN	P37736	vibrio angu	252	6	1.2	211	1	END3_HAEIN	P44319	haemophilus

253	6	1.2	211	1	HI2_MOUSE	P15864	mus musculu	326	6	1.2	268	1	TRPA_PSEAE	P07344	pseudomonas
254	6	1.2	211	1	UREG_ACPPL	O54424	actinobacil	327	6	1.2	269	1	BASI_HUMAN	P35631	h basigin p
255	6	1.2	211	1	URK_BACDH	O9Kd8	bacillus ha	328	6	1.2	269	1	TRPA_PSEPU	P11081	pseudomonas
256	6	1.2	212	1	ATPO_YEAST	P09457	saccharomyc	329	6	1.2	269	1	TRPA_PSESY	P34816	pseudomonas
257	6	1.2	212	1	YEAS_ECOLI	P76249	escherichia	330	6	1.2	269	1	TSNR_STRCN	P18644	streptomyces
258	6	1.2	213	1	HI3_FABIT	P02251	oryctolagus	331	6	1.2	270	1	Y041_CHLTR	O84045	chlamydia t
259	6	1.2	213	1	RB4A_HUMAN	P20338	homo sapien	332	6	1.2	270	1	Y382_RHIME	P25885	thizobium m
260	6	1.2	213	1	RB4A_MOUSE	P56371	mus musculu	333	6	1.2	274	1	GADY_ECO57	P58820	escherichia
261	6	1.2	213	1	RB4A_RAT	P05714	rattus norv	334	6	1.2	274	1	GADY_ECOLI	P37639	escherichia
262	6	1.2	213	1	RB4B_HUMAN	P22750	homo sapien	335	6	1.2	274	1	NHS_STRAS	P52391	streptomyces
263	6	1.2	213	1	RB4B_RAT	P51146	rattus norv	336	6	1.2	275	1	PSBO_CYAAS	O92611	cyanothece
264	6	1.2	213	1	YLXF_BACSU	P23454	bacillus su	337	6	1.2	276	1	RL2_BACDH	O92911	bacillus ha
265	6	1.2	214	1	PYRE_PASMU	O9CJW4	pasteurella	338	6	1.2	277	1	KSGA_BURSP	O9red9	burkholderi
266	6	1.2	214	1	UREG_BORBR	O06709	bordetella	339	6	1.2	279	1	TRPA_RHIME	O92c8	rhizobium m
267	6	1.2	216	1	IF6_THEAC	O9hm6	thermoplasma	340	6	1.2	281	1	PANC_XYIFA	O9pgt8	xyliella fas
268	6	1.2	217	1	MMCI_METAC	P58977	methanosarc	341	6	1.2	283	1	GSPK_XANCP	P34056	xanthomonas
269	6	1.2	217	1	MMCI_METMA	P58977	methanosarc	342	6	1.2	283	1	SURE_CHLNU	O9pk44	chlamydia m
270	6	1.2	217	1	MMCI_METMA	P58977	methanosarc	343	6	1.2	283	1	SURE_CHLTR	O84220	chlamydia t
271	6	1.2	218	1	IF6_THEVO	P08284	gallus gall	344	6	1.2	284	1	HSF4_ARATH	O96320	arabidopsis
272	6	1.2	218	1	IF6_THEVO	O97cud	thermoplasma	345	6	1.2	284	1	YAGM_ECOLI	P71296	escherichia
273	6	1.2	218	1	YDB2_SCHPO	P58978	methanosarc	346	6	1.2	285	1	STAR_BOVIN	O28918	bos taurus
274	6	1.2	219	1	YDB2_SCHPO	O10355	schizosacch	347	6	1.2	285	1	STAR_SHEEP	P79245	ovis aries
275	6	1.2	220	1	RNH2_BRIME	O68821	brucella me	348	6	1.2	287	1	PDX_ECOLI	P77150	escherichia
276	6	1.2	221	1	TRPF_HALVO	P52553	halobacteri	349	6	1.2	288	1	PDX_HAELN	P44630	haemophilus
277	6	1.2	223	1	RB4A_ECHMU	O9gpj3	echinococcu	350	6	1.2	289	1	HTPX_HAELN	O9h92	halobacteri
278	6	1.2	223	1	Y838_BACSU	O58248	bacillus su	351	6	1.2	289	1	PDX_PROMI	O51892	proteus mir
279	6	1.2	223	1	Y838_METUA	O58248	methanococc	352	6	1.2	292	1	Y152_HUMAN	O14152	homo sapien
280	6	1.2	223	1	YUGA_CAEEL	P1851	caenorhabdi	353	6	1.2	294	1	EFTS_LISNO	O92b02	listeria in
281	6	1.2	225	1	UREG_HAELN	P44396	haemophilus	354	6	1.2	294	1	EFTS_LISNO	O8467	listeria mo
282	6	1.2	226	1	NRF8_HAELN	P45016	haemophilus	355	6	1.2	294	1	KC2B_CANAL	O59906	candida alb
283	6	1.2	227	1	RNH2_AERPE	O9yel5	aeropyrum p	356	6	1.2	296	1	CEBB_MOUSE	P28033	mus musculu
284	6	1.2	230	1	PUR7_THEMA	O8xxx9	ralesonia s	357	6	1.2	296	1	CRK_XENLA	P30011	escherichia
285	6	1.2	230	1	TRPF_PALSO	P58321	mus musculu	358	6	1.2	296	1	NADC_ECOLI	P77938	rhodospirill
286	6	1.2	233	1	UBI4_MOUSE	O58858	methanococc	359	6	1.2	296	1	NADC_RHORI	P40380	bacillus su
287	6	1.2	233	1	YEB3_METUA	P50376	costaria co	360	6	1.2	296	1	NADC_SALTY	P30012	salmonella
288	6	1.2	235	1	EFTU_COSCS	P50063	gloeotheca	361	6	1.2	296	1	URER_ECOLI	P32336	escherichia
289	6	1.2	235	1	EFTU_PLEBO	O56332	paracoccusa	362	6	1.2	297	1	HSLO_PSEAE	O9ht26	pseudomonas
290	6	1.2	237	1	NAPC_PARN	O56332	paracoccusa	363	6	1.2	298	1	HMGL_CHICK	P35951	gallus gall
291	6	1.2	237	1	NAPC_PARN	O9X017	thermotoga	364	6	1.2	298	1	Y535_UREPA	O9ppv5	ureaplasma
292	6	1.2	238	1	RNH2_THEMA	P22405	streptomyces	365	6	1.2	300	1	MTG9_CITFR	P40380	bacillus su
293	6	1.2	241	1	KORA_STRLI	O98987	serinus can	366	6	1.2	300	1	MTX1_XANCC	P14243	citrobacter
294	6	1.2	241	1	NEUM_SERCA	P17319	herpes simp	367	6	1.2	300	1	SYGA_PASMU	P30774	xanthomonas
295	6	1.2	245	1	ICP3_HSV1N	P42216	escherichia	368	6	1.2	301	1	SYGA_PASMU	P58794	pasteurella
296	6	1.2	245	1	KSUS_ECOLI	P38774	saccharomyc	369	6	1.2	301	1	ERA_ECO57	P06616	escherichia
297	6	1.2	246	1	DOG2_YEAST	P38774	saccharomyc	370	6	1.2	301	1	ERA_ECOLI	O82445	salmonella
298	6	1.2	246	1	DOG2_YEAST	P38774	saccharomyc	371	6	1.2	301	1	ERA_SALTY	O82445	salmonella
299	6	1.2	246	1	PSA5_TRIBB	O9XZ55	trypanosoma	372	6	1.2	301	1	ERA_SALTY	O56037	salmonella
300	6	1.2	247	1	FAG1_SYNY3	P73574	synechocyst	373	6	1.2	301	1	SYGA_NEIMA	O9jxc6	neisseria m
301	6	1.2	247	1	TFAD_ECOLI	P77699	escherichia	374	6	1.2	301	1	VNSM_TSMVI	O93ic2	tomato spot
302	6	1.2	248	1	ICP3_HSV11	P36313	herpes simp	375	6	1.2	302	1	CRTE_PANAN	P21684	pancoea ana
303	6	1.2	250	1	EXR1_ARATH	O23547	arabidopsis	376	6	1.2	302	1	ERA_HAELN	P43788	haemophilus
304	6	1.2	250	1	UREF_UREPA	O56560	ureaplasma	377	6	1.2	302	1	PTB_CLOBE	O05624	clostridium
305	6	1.2	251	1	A32B_HUMAN	O92668	homo sapien	378	6	1.2	302	1	SYGA_HAELN	P43821	haemophilus
306	6	1.2	251	1	RS6_ASPOF	O9m338	asparagus o	379	6	1.2	303	1	CRKL_HUMAN	P46109	homo sapien
307	6	1.2	251	1	TPIS_PSEAE	O9h5v1	pseudomonas	380	6	1.2	303	1	CRKL_MOUSE	P47941	mus musculu
308	6	1.2	251	1	TPIS_PSESY	P95576	pseudomonas	381	6	1.2	303	1	ERA_YERPE	O82d71	yersinia pe
309	6	1.2	252	1	ICP3_HSV1D	P37318	herpes simp	382	6	1.2	303	1	FMT_RICCN	O33519	ricketsia
310	6	1.2	252	1	PMG2_HUMAN	P15259	homo sapien	383	6	1.2	303	1	VNSM_INSV	O01268	impatiens n
311	6	1.2	253	1	CB2_DUNTE	P27517	dunaliella	384	6	1.2	304	1	CRK_HUMAN	P46108	homo sapien
312	6	1.2	253	1	NCAP_TOSV	P21701	toscana vir	385	6	1.2	304	1	CRK_MOUSE	O64010	mus musculu
313	6	1.2	253	1	TATC_AZOCN	P54085	azotobacter	386	6	1.2	304	1	CRK_MOUSE	O63788	rattus norv
314	6	1.2	255	1	YPMR_BACSU	P40766	bacillus su	387	6	1.2	305	1	Y007_MYCTU	P71575	mycobacteri
315	6	1.2	256	1	CB2_CHIMO	P22666	chlamydomon	388	6	1.2	305	1	YEH2_ECOLI	P33362	escherichia
316	6	1.2	258	1	RNH2_LACIA	O9c917	laetococcus	389	6	1.2	306	1	ERA_PASMU	O9cpn8	pasteurella
317	6	1.2	258	1	SURE_CANVE	O9p1k6	campylobact	390	6	1.2	306	1	HUT6_HALNI	O9hgq7	halobacteri
318	6	1.2	258	1	YDF2_SCHPO	O10474	schizosacch	391	6	1.2	308	1	FMT_NEIMA	O9jwy9	neisseria m
319	6	1.2	259	1	DEFM_ARATH	O9fvs3	arabidopsis	392	6	1.2	309	1	PDXK_ARATH	O8w1k2	arabidopsis
320	6	1.2	259	1	TRUA_METUA	O59069	methanococc	393	6	1.2	310	1	MCAL_MOUSE	P31230	mus musculu
321	6	1.2	261	1	ICP3_HSV1F	O9nrei	homo sapien	394	6	1.2	310	1	BLAC_YEAST	O03082	saccharomyc
322	6	1.2	263	1	ICP3_HSV1F	P08353	herpes simp	395	6	1.2	310	1	BLAC_STRAU	P10509	streptomyces
323	6	1.2	264	1	LDLR_BOVIN	P01131	bos taurus	396	6	1.2	311	1	MSAB_STRGC	O91am9	s peptide m
324	6	1.2	267	1	AXIT_ARATH	O24408	arabidopsis	397	6	1.2	311	1	PDXK_HUMAN	O00764	homo sapien
325	6	1.2	267	1	Y123_THEMA	O9wxk7	thermotoga	398	6	1.2	312	1			

399	6	1.2	312	1	PDXK_SHEEP	P82197	ovis aries	472	6	1.2	359	1	YKHS_CABEL	P34273	caenorhabdi
400	6	1.2	312	1	Y426_ARCFU	029823	archaeoglob	473	6	1.2	360	1	MOA2_MYCTU	053881	mycobacteri
401	6	1.2	312	1	YEC9_YEAST	P39988	saccharomyc	474	6	1.2	360	1	Y303_HELPU	092m03	helicobacte
402	6	1.2	315	1	YNEF_ECO57	08x324	escherichia	475	6	1.2	360	1	Y303_HELPU	025074	helicobacte
403	6	1.2	315	1	YNEF_ECOLI	P76147	escherichia	476	6	1.2	361	1	COBT_MYCTU	010396	mycobacteri
404	6	1.2	316	1	YH58_AQUAE	067638	aquifex aeo	477	6	1.2	361	1	LIG_PHLA	P20010	phlebia rad
405	6	1.2	316	1	YK95_MYCTU	Q10704	mycobacteri	478	6	1.2	362	1	YBDB_ECOLI	P45579	escherichia
406	6	1.2	317	1	YN8F_YEAST	P53727	saccharomyc	479	6	1.2	363	1	ALF8_HUMAN	P05052	homo sapien
407	6	1.2	319	1	TRXB_BUCAP	P57399	buchnera ap	480	6	1.2	364	1	VA0D_NEUCR	P53659	neurospora
408	6	1.2	319	1	TRXB_BUCAP	P81433	buchnera ap	481	6	1.2	364	1	YM31_MYCTU	010503	mycobacteri
409	6	1.2	321	1	PDXK_CABEL	001824	caenorhabdi	482	6	1.2	365	1	FLC1_PROMI	P42272	proteus mir
410	6	1.2	324	1	GPR_THENT	Q81877	thermoanaer	483	6	1.2	365	1	RF2_LACLA	09G9X1	lactococcus
411	6	1.2	324	1	YK95_MYCLE	P54075	mycobacteri	484	6	1.2	365	1	SUCC_METTH	027115	methanobact
412	6	1.2	325	1	HMGL_RAT	P97519	rattus norv	485	6	1.2	365	1	YM27_YEAST	P28652	saccharomyc
413	6	1.2	325	1	SECF_HAEIN	P45590	haemophilus	486	6	1.2	366	1	IC13_HUMAN	P30506	homo sapien
414	6	1.2	327	1	NIFH_CYAA5	007641	cyanothece	487	6	1.2	366	1	VUL_HSV6U	001349	human herpe
415	6	1.2	329	1	COA4_STRCO	086779	streptomyce	488	6	1.2	366	1	YVAF_BACSU	P37518	bacillus su
416	6	1.2	329	1	Y493_MYCTU	Q11158	mycobacteri	489	6	1.2	368	1	SSR2_BOVIN	P34993	bos taurus
417	6	1.2	331	1	PS02_ARATH	Q98841	arabidopsis	490	6	1.2	368	1	Y812_AQUAE	066994	aquifex aeo
418	6	1.2	331	1	SP2B_BACSU	P37575	bacillus su	491	6	1.2	369	1	FORL_BUCAP	08K918	buchnera ap
419	6	1.2	333	1	ILVC_MYCAV	059500	mycobacteri	492	6	1.2	370	1	DNAJ_BACSD	09Kd71	bacillus ha
420	6	1.2	333	1	ILVC_MYCLE	033114	mycobacteri	493	6	1.2	370	1	RECF_BACSU	P05651	bacillus su
421	6	1.2	333	1	ILVC_MYCTU	053248	mycobacteri	494	6	1.2	371	1	ISDF_CAMJB	09pm68	c isepd/ispf
422	6	1.2	334	1	FRUR_ECOLI	P21168	escherichia	495	6	1.2	372	1	FLGI_XANAC	08P127	xanthomonas
423	6	1.2	334	1	FRUR_SALTY	P21930	salmonella	496	6	1.2	372	1	FLGI_XANCP	08P960	xanthomonas
424	6	1.2	334	1	RVUB_BACSU	032055	bacillus su	497	6	1.2	372	1	OPRD_HUMAN	P33130	homo sapien
425	6	1.2	336	1	BLAD_METTH	027717	methanobact	498	6	1.2	372	1	OPRD_MOUSE	P33533	rattus norv
426	6	1.2	337	1	COBU_PSEDE	P29935	pseudomonas	499	6	1.2	372	1	OPRD_RAT	P33533	rattus norv
427	6	1.2	337	1	GPI1_GINLA	P53429	glardia lam	500	6	1.2	375	1	GCST_ARPE	09Yb22	aeropyrum p
428	6	1.2	337	1	ILVC_COREF	Q81P41	corynebacte	501	6	1.2	376	1	G4P1_YEAST	P46672	saccharomyc
429	6	1.2	337	1	YF02_CLOPE	P58668	clostridium	502	6	1.2	376	1	NIRK_RHIME	Q92229	rhizobium m
430	6	1.2	338	1	DCUP_AQUAE	066667	aquifex aeo	503	6	1.2	377	1	DNAJ_BRABA	P94319	bradyrhizob
431	6	1.2	340	1	ILVC_COREG	057179	corynebacte	504	6	1.2	377	1	GP27_RAT	09Jfj3	rattus norv
432	6	1.2	340	1	DCUP_CAMLE	09pn54	campylobact	505	6	1.2	377	1	RNGI_HUMAN	006587	homo sapien
433	6	1.2	340	1	MYB1_MAIZE	P20024	zea mays (m	506	6	1.2	378	1	ELVA_BACSP	P20724	bacillus sp
434	6	1.2	340	1	RECA_THETH	P48297	thermus the	507	6	1.2	378	1	MOBL_THIFE	P20085	thiobacillu
435	6	1.2	341	1	CYTR_ECOLI	P06964	escherichia	508	6	1.2	379	1	GP27_MOUSE	P20085	mus muscullu
436	6	1.2	342	1	COBT_METSO	Q9x7E4	methylobact	509	6	1.2	379	1	NIR_RHIGA	001537	rhizobium g
437	6	1.2	343	1	DHSO_BACHD	Q929U1	bacillus ha	510	6	1.2	380	1	DNAJ_BACST	045552	bacillus st
438	6	1.2	343	1	HMD_METVO	Q50840	methanococc	511	6	1.2	382	1	GRK_BACSU	P42100	bacillus su
439	6	1.2	343	1	RADA_HAILO	Q88328	halodacteri	512	6	1.2	382	1	KCCI_METAN	014408	metathium
440	6	1.2	344	1	HMD_METTF	Q50758	halodacteri	513	6	1.2	385	1	SNFS_HUMAN	Q12824	homo sapien
441	6	1.2	344	1	HMD_METTH	Q27211	m coenzyme	514	6	1.2	385	1	SNFS_MOUSE	Q920H3	mus muscullu
442	6	1.2	344	1	HMD_METTH	P32440	m coenzyme	515	6	1.2	386	1	YBDL_ECOLI	P7780L	escherichia
443	6	1.2	344	1	HMD_METTH	P81221	m coenzyme	516	6	1.2	388	1	NORA_STRAA	P21311	staphylococ
444	6	1.2	344	1	HMD_METTH	Q10078	echinosacch	517	6	1.2	390	1	SUCC_COXBU	P53592	coxiella bu
445	6	1.2	345	1	HRCA_LISIN	Q922B6	listeria in	518	6	1.2	392	1	G3PA_TOBAC	P09043	nicotiana t
446	6	1.2	345	1	HRCA_LISIMO	Q95566	listeria mo	519	6	1.2	392	1	VATC_YEAST	P31412	saccharomyc
447	6	1.2	346	1	FT22_RHIME	P45484	rhizobium m	520	6	1.2	393	1	CYB_METSE	Q47489	metridium s
448	6	1.2	347	1	PLIG_AGR75	Q44458	agrobacteri	521	6	1.2	393	1	YE60_YEAST	P32611	saccharomyc
449	6	1.2	347	1	PUR7_HAINT	Q9hnU7	halobacteri	522	6	1.2	398	1	FADH_PSEPU	P46134	pseudomonas
450	6	1.2	347	1	Y576_METJA	057996	halobacteri	523	6	1.2	398	1	HRCA_CHLPP	Q92850	chlamydia p
451	6	1.2	347	1	Y576_METJA	057996	methanococc	524	6	1.2	398	1	MPK2_CHICK	090881	gallus galli
452	6	1.2	348	1	AROH_ERWHE	P34583	caenorhabdi	525	6	1.2	398	1	NIS2_ANAVA	Q44482	anabena va
453	6	1.2	348	1	RTCA_ANASP	Q8Y459	erwinia her	526	6	1.2	399	1	EFTU_CAMJE	069393	campylobact
454	6	1.2	350	1	Y098_SYNY3	Q58863	anabena sp	527	6	1.2	400	1	CN37_BOVIN	P06633	bos taurus
455	6	1.2	351	1	NOEL_RHIFR	085713	rhizobium f	528	6	1.2	400	1	MPK2_HUMAN	P36507	homo sapien
456	6	1.2	351	1	NOEL_RHIFR	P53534	rhizobium s	529	6	1.2	400	1	MPK2_RAT	P36506	rattus norv
457	6	1.2	353	1	GAG_METVG	P03343	mouse mamma	530	6	1.2	400	1	PGK_CAMJE	Q9pm65	campylobact
458	6	1.2	353	1	HMD_METTL	Q50759	m coenzyme	531	6	1.2	400	1	PUR2_THBMA	Q9X0X7	thermotoga
459	6	1.2	354	1	E2B_XYLPA	Q9pa55	xyliella fas	532	6	1.2	401	1	MPK2_MOUSE	063902	mus muscullu
460	6	1.2	354	1	MURG_ECO57	08x9V8	escherichia	533	6	1.2	402	1	ARGJ_METJA	057663	methanococc
461	6	1.2	354	1	MURG_ECO57	08E164	escherichia	534	6	1.2	402	1	CSD_UREPA	Q9pq35	ureaplaasma
462	6	1.2	354	1	MURG_ECOLI	P17443	escherichia	535	6	1.2	402	1	FIXF_RHISN	P55467	rhizobium s
463	6	1.2	354	1	YBPI_YEAST	P38616	saccharomyc	536	6	1.2	403	1	ARGJ_CAMJC	Q9a3J4	c arginine
464	6	1.2	355	1	GBAC_CABEL	Q19572	caenorhabdi	537	6	1.2	404	1	HMDH_HALVO	059494	halobacteri
465	6	1.2	355	1	MURG_NEIMA	Q9jsz7	neisseria m	538	6	1.2	404	1	CGE2_HUMAN	096020	homo sapien
466	6	1.2	355	1	MURG_NEIMB	Q9K0Y2	neisseria m	539	6	1.2	404	1	CGE2_MOUSE	092228	mus muscullu
467	6	1.2	356	1	Y565_ASRF7	P33165	african swi	540	6	1.2	404	1	HISX_ARCFU	057404	archaeoglob
468	6	1.2	356	1	YD6B_SCHPO	Q10323	schizosacch	541	6	1.2	404	1	Y867_HAEIN	057484	haemophilus
469	6	1.2	357	1	KARG_ERISI	Q9nh48	eritochel s	542	6	1.2	405	1	TRPB_XYLPA	09pd44	xyliella fas
470	6	1.2	358	1	HMD_METJA	Q58194	m coenzyme	543	6	1.2	405	1	YLPF_YEAST	P38746	saccharomyc
471	6	1.2	358	1	PONI_RABIT	P27170	oryctolagus	544	6	1.2	408	1	APGI_ARCFU	028523	archaeoglob

545	6	1.2	409	1	EFTU_ASTLO	P14614	astasia lon	618	6	1.2	463	1	ZABA_SCHPO	Q12702	schizosacch
546	6	1.2	409	1	EFTU_CVAPA	P17245	cyamophora	619	6	1.2	463	1	YIBO_HAEIN	P44903	haemophilus
547	6	1.2	409	1	EFTU_EUGGR	P02991	euglena gra	620	6	1.2	464	1	FLID_AERPU	Q91976	aeromonas p
548	6	1.2	410	1	EFTU_CVACA	O9c1v8	cyanidium c	621	6	1.2	464	1	YTI3_CAEBL	Q10917	caenorhabdi
549	6	1.2	412	1	CLPX_AOUAE	O67156	agulfex aeo	622	6	1.2	465	1	SNX8_HUMAN	Q9Y552	homo sapien
550	6	1.2	412	1	DAD3_RHILQ	O981x2	rhizobium 1	623	6	1.2	467	1	MATK_RAT	P41243	rattus norv
551	6	1.2	412	1	DJAD2_HUMAN	O60864	homo sapien	624	6	1.2	467	1	NX1B_BOVIN	Q28142	bos taurus
552	6	1.2	412	1	DJAD2_MOUSE	O9qy10	mus musculu	625	6	1.2	467	1	SYC_THETN	O87473	thermomae
553	6	1.2	412	1	DJAD2_RAT	O35824	rattus norv	626	6	1.2	468	1	6PGB_BACSU	P12013	bacillus su
554	6	1.2	413	1	PBP2_YEAST	P38151	saccharomyc	627	6	1.2	468	1	NX1B_RAT	O63373	rattus norv
555	6	1.2	413	1	PROA_CAICR	O9a2x6	caulobacter	628	6	1.2	469	1	CHLN_SYNY3	P28172	synchoeyst
556	6	1.2	414	1	CPXA_PSEPU	P00183	pseudomonas	629	6	1.2	469	1	TRPC_PASMU	P57855	pasteurella
557	6	1.2	414	1	KCCI_EMENT	O00771	emeticella	630	6	1.2	470	1	NR54_HUMAN	Q15233	homo sapien
558	6	1.2	415	1	PAX2_MOUSE	P32114	mus musculu	631	6	1.2	471	1	MURD_STRCO	O9s249	streptomyce
559	6	1.2	416	1	PAX2_HUMAN	O02962	homo sapien	632	6	1.2	472	1	KPYK_MYCTO	O06134	mycobacteri
560	6	1.2	416	1	RAGE_BOVIN	O28173	bos taurus	633	6	1.2	472	1	PLS8_SPIOL	Q43689	spinae o1
561	6	1.2	417	1	GLVA_BORBU	O51547	borrelia bu	634	6	1.2	474	1	NUAM_PARTE	P15561	paramecium
562	6	1.2	417	1	BGK_CANMA	P41757	candida mal	635	6	1.2	476	1	EFTU_ARATH	P17745	arabidopsis
563	6	1.2	418	1	PROA_CIOAB	O97662	clostridium	636	6	1.2	476	1	THI1_THEVO	Q97466	thermoplas
564	6	1.2	419	1	CLPX_CHLTR	O84711	chlamydia t	637	6	1.2	477	1	MURC_XANAC	O8p37	xanthomonas
565	6	1.2	420	1	CN37_MOUSE	P16310	mus musculu	638	6	1.2	477	1	MURC_XANCP	O8pc39	xanthomonas
566	6	1.2	421	1	CN37_HUMAN	P09543	homo sapien	639	6	1.2	477	1	MURC_XYLFA	O9f680	xyella fas
567	6	1.2	422	1	ADHF_SCHPO	O09669	schizosacch	640	6	1.2	478	1	HRTG_CROAT	P34182	crotaeus at
568	6	1.2	424	1	CLPX_RALSO	O8xyp6	raistonia s	641	6	1.2	478	1	MAOX_BACST	P16468	bacillus st
569	6	1.2	424	1	PUR2_HELPY	O9zk07	helicobacte	642	6	1.2	478	1	T23C_BACTB	P12250	bacillus th
570	6	1.2	424	1	PUR2_HELPY	O25817	helicobacte	643	6	1.2	479	1	EFTI_SOYAN	Q43467	glycine max
571	6	1.2	424	1	VID2_AGRTU	P06668	agrobacteri	644	6	1.2	479	1	P2CB_HUMAN	O75668	homo sapien
572	6	1.2	425	1	SYS_STRCO	O9zbx1	streptomyce	645	6	1.2	480	1	NR61_HUMAN	O15405	homo sapien
573	6	1.2	426	1	PROA_RALSO	O8xvt6	raistonia s	646	6	1.2	482	1	BHB3_HUMAN	O9c0j9	homo sapien
574	6	1.2	426	1	STZ5_HUMAN	O00506	homo sapien	647	6	1.2	483	1	MURC_RALSO	O8xv18	raistonia s
575	6	1.2	426	1	STZ5_MOUSE	O9z2w1	mus musculu	648	6	1.2	483	1	TRKH_ECOLI	P21166	escherichia
576	6	1.2	427	1	HISX_STRMU	O8dtq7	streptococc	649	6	1.2	483	1	TRKH_ECOLI	Q91616	escherichia
577	6	1.2	427	1	TIG_LACIA	O9c1i5	lactococcu	650	6	1.2	484	1	THI1_PSEAE	O9s563	pseudomonas
578	6	1.2	427	1	TIG_STRPN	O97669	streptococc	651	6	1.2	484	1	UDPE_NPVOP	O65163	oryzias pseu
579	6	1.2	428	1	ENPE_HUMAN	O73356	homo sapien	652	6	1.2	491	1	YOIE_CAEBL	P34629	caenorhabdi
580	6	1.2	428	1	PROA_STRCO	O97dxi	scieptomyce	653	6	1.2	492	1	FSCT_HUMAN	Q16658	homo sapien
581	6	1.2	429	1	COBB_RHIME	O92p48	rhizobium m	654	6	1.2	492	1	FSCT_MOUSE	Q61553	mus musculu
582	6	1.2	429	1	GSAB_BACSU	P71084	bacillus su	655	6	1.2	493	1	DCMC_HUMAN	O9s822	homo sapien
583	6	1.2	429	1	LAMB_KLEPN	P31242	klebsiella	656	6	1.2	494	1	NORV_VIBVU	O8d4f8	vibrio vuln
584	6	1.2	430	1	NY42_YEAST	P49686	saccharomyc	657	6	1.2	495	1	NR61_MOUSE	O64239	mus musculu
585	6	1.2	430	1	SYMC_YEAST	Q12109	saccharomyc	658	6	1.2	495	1	TRME_TREPA	O83561	treponema p
586	6	1.2	433	1	COBB_PSEDE	P21632	pseudomonas	659	6	1.2	497	1	AMPA_RHIME	Q92qy7	rhizobium p
587	6	1.2	434	1	CISY_BRAJA	-P94335	bradyrhizob	660	6	1.2	498	1	DGTL_RAT	O9emt3	rattus norv
588	6	1.2	435	1	PURA_CHLTE	O8kak6	chlocriobium	661	6	1.2	498	1	PRIM_AOUAE	O67465	agulfex aeo
589	6	1.2	436	1	Y868_CHLMU	O9p3j1	chlamydia m	662	6	1.2	498	1	VGLX_PRIVT	P07562	pseudorabie
590	6	1.2	437	1	CTBO_XENTLA	O9w758	xenopus lae	663	6	1.2	499	1	CIW5_HUMAN	O95279	homo sapien
591	6	1.2	439	1	AROA_DEIRA	O9rvd3	deinococcus	664	6	1.2	499	1	CPNI_RAT	P13393	rattus norv
592	6	1.2	440	1	BCHN_HEIMO	O92969	heliobacilli	665	6	1.2	500	1	ICI_HUMAN	P05155	homo sapien
593	6	1.2	440	1	GAGC_AYISC	P05433	avian sarco	666	6	1.2	501	1	PYCA_METJA	O58526	methanococc
594	6	1.2	440	1	YHJE_ECOLI	P37643	escherichia	667	6	1.2	502	1	SPHM_HUMAN	P51688	homo sapien
595	6	1.2	442	1	NX1B_HUMAN	P58400	homo sapien	668	6	1.2	503	1	AMV8_WHEAT	P93594	tritlicum ae
596	6	1.2	443	1	MT17_YEAST	P06106	saccharomyc	669	6	1.2	503	1	FLUT_DROME	O9vui9	drosophila
597	6	1.2	443	1	MURC_LACIA	O9ce10	lactococcu	670	6	1.2	503	1	NR61_HUMAN	Q16656	homo sapien
598	6	1.2	446	1	KCCI_YEAST	P27466	saccharomyc	671	6	1.2	503	1	NR61_MOUSE	Q9w400	mus musculu
599	6	1.2	447	1	KCC2_YEAST	P22517	saccharomyc	672	6	1.2	504	1	FLIC_SALBU	O06690	salmonella
600	6	1.2	448	1	PO31_HUMAN	Q03052	homo sapien	673	6	1.2	504	1	FLIC_SALDE	O06970	salmonella
601	6	1.2	449	1	LAT_MYCTU	P96895	mycobacteri	674	6	1.2	504	1	FLIC_SALDU	O06971	salmonella
602	6	1.2	449	1	PO31_MOUSE	P21952	mus musculu	675	6	1.2	504	1	FLIC_SALBN	O06972	salmonella
603	6	1.2	450	1	LAT_NOCIA	O05174	nocardia la	676	6	1.2	504	1	FLIC_SALWC	O06981	salmonella
604	6	1.2	451	1	PO31_RAT	P20267	rattus norv	677	6	1.2	504	1	FLIC_SALMO	O06973	salmonella
605	6	1.2	451	1	YOEV_BACSU	P54462	bacillus su	678	6	1.2	504	1	FLIC_SALNA	O06974	salmonella
606	6	1.2	452	1	YP59_MYCTU	O50739	mycobacteri	679	6	1.2	504	1	FLIC_SALRO	O06982	salmonella
607	6	1.2	454	1	RHBE_RHIME	O923g8	rhizobium m	680	6	1.2	504	1	FLIC_SALSE	O06983	salmonella
608	6	1.2	454	1	UCR2_NEUCR	O60044	neutrospora	681	6	1.2	504	1	ICI_MOUSE	P97290	mus musculu
609	6	1.2	456	1	MANE_SALMO	O01411	salmonella	682	6	1.2	505	1	NR61_BEAR	Q90x44	brachydanio
610	6	1.2	456	1	MANB_SALTY	P26341	salmonella	683	6	1.2	505	1	MARK_MOUSE	P41242	mus musculu
611	6	1.2	456	1	BPXA_BOVIN	O28143	bos taurus	684	6	1.2	506	1	MURE_SYNY3	O05469	synchoeyst
612	6	1.2	457	1	BPXA_COWTE	O46372	comamonas t	685	6	1.2	505	1	PALA_PRAVU	P07218	phaseolus v
613	6	1.2	457	1	LAT_STRCL	O01767	streptomyce	686	6	1.2	507	1	C7C3_ARATH	O06468	salmonella
614	6	1.2	457	1	Y976_TREPA	O83941	treponema p	687	6	1.2	507	1	FLIC_SALBE	O06974	salmonella
615	6	1.2	460	1	EFIA_NEUCR	O01372	neutrospora	688	6	1.2	507	1	FLIC_SALON	O06974	salmonella
616	6	1.2	460	1	HEMO_MOUSE	O91x72	mus musculu	689	6	1.2	507	1	MATK_HUMAN	P42679	homo sapien
617	6	1.2	460	1	HEMO_RAIT	P20058	oryctolagus	690	6	1.2	508	1	PSBB_EUGGR	P14813	euglena gra

691	6	1.2	508	1	RPB2_THEAC	Q03586 thermoplasm	764	6	1.2	576	1	MOES_HUMAN	P26038 homo sapien
692	6	1.2	509	1	GUAA_CLOPE	Q8146 clostridium	765	6	1.2	576	1	MOES_MOUSE	P26041 mus musculu
693	6	1.2	510	1	BWP6_MOUSE	P20722 mus musculu	766	6	1.2	576	1	MOES_PIG	P26042 sus scrofa
694	6	1.2	510	1	GPMT_PSES	P52813 pseudomonas	767	6	1.2	576	1	SBP_CARBL	Q21950 caenorhabdi
695	6	1.2	511	1	AROF_LYCES	P37215 lycopericon	768	6	1.2	577	1	SYR_VIBCH	Q91965 vibrio chol
696	6	1.2	511	1	AROG_SOLTU	P37822 solanum tub	769	6	1.2	578	1	DSBD_PSESP	Q9KJ23 pseudomonas
697	6	1.2	511	1	COPD_BOVIN	P53619 bos taurus	770	6	1.2	579	1	SYQ_XANAC	Q8pn35 xanthomonas
698	6	1.2	511	1	COPD_HUMAN	P48444 homo sapien	771	6	1.2	579	1	SYQ_XANAC	Q8pn35 xanthomonas
699	6	1.2	512	1	C26A_HUMAN	Q9n63 homo sapien	772	6	1.2	580	1	DAKI_SCHPO	O13992 schizosacch
700	6	1.2	513	1	HUTH_RHIL	Q9810 rhizobium 1	773	6	1.2	580	1	EZRI_BOVIN	P31976 bos taurus
701	6	1.2	514	1	VG64_HSV1	Q00156 ictalurid h	774	6	1.2	580	1	MEND_BACSU	P23190 b menaquin
702	6	1.2	516	1	BOCT_RAT	Q9P290 ratius norv	775	6	1.2	583	1	KPYA_RICCO	Q43117 ricinus com
703	6	1.2	516	1	CE23_HORVU	P52711 hordeum vul	776	6	1.2	583	1	RADI_HUMAN	P35241 homo sapien
704	6	1.2	522	1	MSAB_NEIGO	P14930 n pepticid m	777	6	1.2	583	1	RADI_MOUSE	P26043 mus musculu
705	6	1.2	522	1	MSAB_NEIMA	Q91mb n pepticid m	778	6	1.2	583	1	RADI_PIG	P26044 sus scrofa
706	6	1.2	522	1	MSAB_NEIMB	Q91mb n pepticid m	779	6	1.2	584	1	YMB3_YEAST	O04228 saccharomyc
707	6	1.2	524	1	GUAA_YEAST	P38625 saccharomyc	780	6	1.2	585	1	EZRI_HUMAN	P15311 homo sapien
708	6	1.2	527	1	TP6B_PYRAE	Q8zvm0 pyrodaculm	781	6	1.2	585	1	EZRI_MOUSE	P26040 mus musculu
709	6	1.2	528	1	FOLE_NEOCR	O13492 neurospora	782	6	1.2	586	1	SYR_STRCO	Q9wx29 streptomyc
710	6	1.2	529	1	NMT_AJECA	P34763 atellomyces	783	6	1.2	587	1	DSO2_PSEAE	Q91104 pseudomonas
711	6	1.2	529	1	NX1A_MOUSE	Q9c884 mus musculu	784	6	1.2	587	1	YNI4_YEAST	P53933 saccharomyc
712	6	1.2	529	1	TIMK_ECOLI	P08957 escherichia	785	6	1.2	588	1	ASH1_YEAST	P34223 saccharomyc
713	6	1.2	529	1	TIM_SALPO	P07989 salmoneilla	786	6	1.2	589	1	STER_CANAL	P46599 candida alb
714	6	1.2	529	1	TIM_SALTY	P40813 salmoneilla	787	6	1.2	591	1	GAG_MMTVB	P10248 mouse mamma
715	6	1.2	530	1	UD12_HUMAN	P36509 homo sapien	788	6	1.2	593	1	KPYA_TOBAC	P40545 nicotiana t
716	6	1.2	530	1	UD17_HUMAN	Q9haw7 homo sapien	789	6	1.2	594	1	DCP2_SCHPO	Q92345 schizosacch
717	6	1.2	530	1	UD18_HUMAN	Q9haw7 homo sapien	790	6	1.2	596	1	FRDA_SHEON	P83223 shewanelia
718	6	1.2	530	1	UD19_HUMAN	O60656 homo sapien	791	6	1.2	596	1	TFB8_YEAST	P29056 saccharomyc
719	6	1.2	530	1	UD1A_HUMAN	Q9haw7 homo sapien	792	6	1.2	599	1	TYPA_HELPJ	Q92103 helicobacte
720	6	1.2	531	1	KEYK_ELITE	O44006 elmeria ten	793	6	1.2	599	1	TYPA_HELPJ	O25225 helicobacte
721	6	1.2	531	1	UD16_HUMAN	P19224 homo sapien	794	6	1.2	600	1	DCP1_ORYSA	Q92755 chlamydia p
722	6	1.2	533	1	UVIN_RHITR	O05467 rhizobium t	795	6	1.2	602	1	PCPI_ORYSA	P51847 oryza sativ
723	6	1.2	533	1	UD11_HUMAN	P22309 homo sapien	796	6	1.2	603	1	RPB1_METTH	O27124 methanobact
724	6	1.2	534	1	K2C4_HUMAN	P19013 homo sapien	797	6	1.2	604	1	RPB1_METTH	P09845 methanobact
725	6	1.2	534	1	UD13_HUMAN	P35503 homo sapien	798	6	1.2	605	1	NUKX_PELISU	O79678 pelomedusa
726	6	1.2	534	1	UD14_HUMAN	P22310 homo sapien	799	6	1.2	607	1	GDST_HUMAN	P52306 homo sapien
727	6	1.2	534	1	UD15_HUMAN	P35504 homo sapien	800	6	1.2	609	1	PPCK_MYCLE	O06084 mycobacteri
728	6	1.2	536	1	CH62_CHLVC	P59698 chlamydophi	801	6	1.2	610	1	ARLC_MAIZE	P13526 zea mays (m
729	6	1.2	538	1	LY41_THHEVA	O9w222 thermotoga	802	6	1.2	611	1	ACFS_FELICA	O62763 felis silve
730	6	1.2	539	1	CH60_FUSNN	O8r573 fusobacteri	803	6	1.2	612	1	ARBS_MAIZE	P13307 zea mays (m
731	6	1.2	539	1	CH60_FUSNP	O8g100 fusobacteri	804	6	1.2	613	1	ACFS_BOVIN	P23795 bos taurus
732	6	1.2	541	1	GT10_HUMAN	O95528 homo sapien	805	6	1.2	614	1	ILVD_RHIL	O98b28 rhizobium 1
733	6	1.2	543	1	CH60_BRECH	Q8r100 brevibacilli	806	6	1.2	619	1	ABS_DROME	O9w360 drosophila
734	6	1.2	544	1	CH63_RHILE	Q91690 rhizobium 1	807	6	1.2	622	1	PTMA_HAEIN	P44092 haemophilus
735	6	1.2	546	1	AAAS_HUMAN	O9n929 homo sapien	808	6	1.2	623	1	Y711_HUMAN	O94813 homo sapien
736	6	1.2	546	1	AAAS_MOUSE	P58743 mus musculu	809	6	1.2	625	1	NIFA_AZOBR	P30657 azospirillum
737	6	1.2	548	1	WR72_ARATH	Q91x98 arabidopsis	810	6	1.2	628	1	LU_HUMAN	P50895 homo sapien
738	6	1.2	552	1	ILVD_CLOAB	O97ee3 clostridium	811	6	1.2	629	1	PABP_ARATH	P42781 arabidopsis
739	6	1.2	554	1	GA44_HUMAN	P48169 homo sapien	812	6	1.2	632	1	PABP_DROME	P21871 drosophila
740	6	1.2	556	1	GCH2_SYNY3	P74104 synechocyst	813	6	1.2	635	1	PTMA_KLEPN	O9xmb7 klebsiella
741	6	1.2	556	1	PT1_STRCO	O9k201 streptomyc	814	6	1.2	637	1	PTMA_ECOLI	P00550 escherichia
742	6	1.2	556	1	SYO_PSEAE	O91248 pseudomonas	815	6	1.2	638	1	GHR_HUMAN	P10912 homo sapien
743	6	1.2	557	1	R81_ECOLI	P02348 escherichia	816	6	1.2	638	1	GHR_MACMU	P79194 macaca mula
744	6	1.2	558	1	GDST_BOVIN	Q04173 bos taurus	817	6	1.2	643	1	CTR2_XENTLA	P79955 xenopus lae
745	6	1.2	559	1	THSA_SULSH	O9v269 sulfolobus	818	6	1.2	643	1	K2C1_HUMAN	P04264 homo sapien
746	6	1.2	560	1	THSA_SULSH	P46129 sulfolobus	819	6	1.2	643	1	SENI_HUMAN	O9p0u3 homo sapien
747	6	1.2	562	1	CHS5_CANAL	O74161 candida alb	820	6	1.2	643	1	Y903_MERYA	O58313 methanococc
748	6	1.2	562	1	DHE3_DROME	P54385 drosophila	821	6	1.2	644	1	HS71_PICAN	P53421 pichia angu
749	6	1.2	563	1	Y288_CHLTR	O84290 chlamydia t	822	6	1.2	644	1	PLB1_SCHPO	O9nrt5 schizosacch
750	6	1.2	565	1	MAOX_SCHPO	P40375 schizosacch	823	6	1.2	644	1	RNR_HELPJ	O9zj39 helicobacte
751	6	1.2	565	1	MOBA_BIFLO	O8gn32 biflobacte	824	6	1.2	648	1	APPC_DICDI	P34059 dictyostella
752	6	1.2	565	1	MOBA_BIFLO	P45283 haemophilus	825	6	1.2	648	1	PRIA_RICPR	O9z2d0 rickettsia
753	6	1.2	566	1	DCPY_HAEIN	P87208 emericella	826	6	1.2	648	1	RCOI_MOUSE	O9z129 mus musculu
754	6	1.2	566	1	PAP_SCHPO	Q10295 schizosacch	827	6	1.2	649	1	RCOI_HUMAN	P46063 homo sapien
755	6	1.2	567	1	GFV_RAT	O08770 ratius norv	828	6	1.2	652	1	NAK1_SCHPO	O75011 schizosacch
756	6	1.2	568	1	OXO_OXAP	P40149 oxalobacter	829	6	1.2	652	1	NOS2_PARDE	O51705 paracoccus
757	6	1.2	569	1	UREI_SYNY3	P73061 synechocyst	830	6	1.2	653	1	API_ACHLY	P15636 achromobact
758	6	1.2	570	1	SMA4_CAREL	P45897 caenorhabdi	831	6	1.2	653	1	PABP_SCHPO	P31209 schizosacch
759	6	1.2	571	1	DISJ_BORTA	P30431 borhrops ja	832	6	1.2	655	1	MLAI_MOUSE	P45700 mus musculu
760	6	1.2	571	1	PAL4_PORPI	Q40910 populus kit	833	6	1.2	660	1	DNLJ_BORBU	O51502 borrelia bu
761	6	1.2	572	1	MAOX_HUMAN	P48163 homo sapien	834	6	1.2	660	1	SYR_HUMAN	P54136 homo sapien
762	6	1.2	572	1	MOES_LYTVA	P53962 lytechinus	835	6	1.2	660	1	SYR_MOUSE	O9d019 mus musculu
763	6	1.2	572	1	UREI_STRSL	P50047 streptococc	836	6	1.2	660	1	YHLI_EBV	P03101 epstein-bar

837	6	1.2	661	1	SVR_CRILLO	P37880	cricketulus	910	6	1.2	766	1	BCSB_SALTY	Q931i1	salmonella
838	6	1.2	662	1	I12R_HUMAN	P42701	homo sapien	911	6	1.2	766	1	GAP1_SCHPO	P33277	schizosacch
839	6	1.2	662	1	LOX2_RAT	Q02759	rattus norv	912	6	1.2	767	1	PSAA_SYSPW	Q976u0	synthosacch
840	6	1.2	662	1	LOXL_MOUSE	P33654	mus musculu	913	6	1.2	773	1	ITB1_BOVIN	P53712	bov taurus
841	6	1.2	662	1	NX2B_RAT	Q63376	rattus norv	914	6	1.2	773	1	PSAA_PROMA	Q914r4	prochloroco
842	6	1.2	663	1	DNLI_CHLTR	Q64148	chlamydia t	915	6	1.2	774	1	RRP3_INCB	P21770	influenza c
843	6	1.2	666	1	NX2B_HUMAN	P58401	homo sapien	916	6	1.2	775	1	TREY_INCB	Q44315	arthrobacte
844	6	1.2	667	1	M121_DROME	P53624	drosophila	917	6	1.2	776	1	TFR1_ARTCH	Q90997	gallus galli
845	6	1.2	667	1	PALY_METUA	Q04058	helianthus	918	6	1.2	776	1	VP4_ROTCH	Q02945	equine rota
846	6	1.2	668	1	PEOB_METUA	Q05196	mechanococc	919	6	1.2	780	1	SAV_SULAC	Q07590	sulfolobus
847	6	1.2	668	1	PAB5_ARATH	Q05196	arabidopsis	920	6	1.2	781	1	MEFV_HUMAN	Q15553	homo sapien
848	6	1.2	668	1	PAPC_BACSU	P42971	bacillus su	921	6	1.2	781	1	PPSA_ECOLI	P23358	escherichia
849	6	1.2	669	1	AMY_ALTHA	P29957	alteromonas	922	6	1.2	781	1	FUR1_MOUSE	P23168	mus musculu
850	6	1.2	669	1	NUSM_ARATH	P23388	arabidopsis	923	6	1.2	785	1	SYFB_VIBVU	Q8da39	vibrio vuln
851	6	1.2	670	1	NUSM_OENBE	P10330	oenothera b	924	6	1.2	797	1	PBPA_NEIRO	Q051j1	neisseria g
852	6	1.2	670	1	NUSM_WHEAT	Q37680	triticum ae	925	6	1.2	797	1	SHK3_HUMAN	Q9bly0	homo sapien
853	6	1.2	671	1	2282_HUMAN	Q9ndv7	homo sapien	926	6	1.2	798	1	ITB1_PELCA	P53373	felis silve
854	6	1.2	675	1	PPK_HELPJ	Q9am10	helicobacte	927	6	1.2	798	1	ITB1_HUMAN	P05556	homo sapien
855	6	1.2	675	1	PPK_HELPJ	Q02654	helicobacte	928	6	1.2	798	1	ITB1_MOUSE	P09055	mus musculu
856	6	1.2	675	1	IVRB_NEIGO	Q05093	neisseria g	929	6	1.2	798	1	PBPA_NEICI	Q86088	neisseria c
857	6	1.2	675	1	IVRB_NEIMA	P56996	neisseria m	930	6	1.2	798	1	PBPA_NEIFL	Q87626	neisseria f
858	6	1.2	675	1	IVRB_NEIMA	Q03395	neisseria m	931	6	1.2	798	1	PBPA_NEIFA	Q87579	neisseria l
859	6	1.2	679	1	HS82_AJECA	P33125	ajellomyces	932	6	1.2	799	1	PBPA_NEIMA	Q051j4	neisseria m
860	6	1.2	679	1	TKET1_YEAST	P23254	saccharomyc	933	6	1.2	799	1	ITB1_RAT	P49314	rattus norv
861	6	1.2	679	1	TKTA_CRAPL	Q42675	craterostig	934	6	1.2	803	1	GYRB_ECOLI	P49314	rattus norv
862	6	1.2	680	1	KALM_HUMAN	P23352	homo sapien	935	6	1.2	803	1	GYRB_SALTY	P06982	escherichia
863	6	1.2	681	1	NUSC_SYNY3	Q05429	synecocyst	936	6	1.2	804	1	MEAG_HUMAN	Q60008	salmonella
864	6	1.2	681	1	VADD_SCHPO	Q09838	schizosacch	937	6	1.2	805	1	L100_ADE02	Q153z0	homo sapien
865	6	1.2	692	1	BOMD_XENLA	P79944	xenopus lae	938	6	1.2	806	1	NUDC_MYCTU	P24932	human adeno
866	6	1.2	702	1	AMYG_CLOSP	P29761	clostridium	939	6	1.2	807	1	L100_ADE05	P95175	mycobacteri
867	6	1.2	703	1	CDGT_BACS2	P31746	bacillus sp	940	6	1.2	808	1	PLSB_VIBPA	Q87kx0	vibrio para
868	6	1.2	703	1	SYM_CHLTE	P53077	chlorobium	941	6	1.2	810	1	KIUS_YEAST	Q39586	chlamydomo
869	6	1.2	705	1	MMLC_STRCO	Q08022	streptomyces	942	6	1.2	814	1	METE_CHLRE	Q13387	homo sapien
870	6	1.2	705	1	KMX_CHICK	Q09059	gallus galli	943	6	1.2	824	1	JIP2_HUMAN	P44323	haemophilus
871	6	1.2	705	1	VKM2_YEAST	P35995	saccharomyc	944	6	1.2	829	1	IFP2_HAERIN	P10799	agrobacteri
872	6	1.2	706	1	UII7_HSVAB	P28950	equine herp	945	6	1.2	829	1	VIRA_AGR79	P07168	agrobacteri
873	6	1.2	707	1	SFPQ_HUMAN	P23246	homo sapien	946	6	1.2	829	1	VIRA_AGR7U	P43572	saccharomyc
874	6	1.2	708	1	PAL2_IPOBA	Q42858	ipomoea bat	947	6	1.2	832	1	YFC4_YEAST	P57873	pasteurella
875	6	1.2	712	1	PAL2_TOBAC	P43515	nicotiana t	948	6	1.2	833	1	IF2_PASMU	P43572	pasteurella
876	6	1.2	712	1	PAL3_TOBAC	P45713	nicotiana t	949	6	1.2	835	1	NIG2_HUMAN	Q13263	homo sapien
877	6	1.2	713	1	PAL1_SOYBN	P27991	glycine max	950	6	1.2	835	1	TF1E_HUMAN	Q13263	homo sapien
878	6	1.2	713	1	SVR_CABEL	Q19825	caenorhabdi	951	6	1.2	836	1	NLG2_RAT	Q62888	rattus norv
879	6	1.2	715	1	PAL1_TOBAC	P25872	nicotiana t	952	6	1.2	843	1	COY_HUMAN	P10643	homo sapien
880	6	1.2	718	1	PAL2_CICAR	Q28mk9	cicer ariet	953	6	1.2	847	1	SYA_HELPJ	Q9zjy5	helicobacte
881	6	1.2	720	1	NUP2_YEAST	P32499	saccharomyc	954	6	1.2	847	1	SYA_HELPJ	P56462	helicobacte
882	6	1.2	720	1	PAL1_SOLTU	P31449	solanum tub	955	6	1.2	851	1	MCEI_FOPMV	Q91584	fowlpox vir
883	6	1.2	721	1	PAL5_LYCDS	P26600	lycopersico	956	6	1.2	853	1	PBPA_HAEIN	P31776	h penicillii
884	6	1.2	722	1	PAL1_CITUL	Q42667	citrus limo	957	6	1.2	861	1	POS8_CABEL	P34552	caenorhabdi
885	6	1.2	723	1	PAL1_PEA	Q04561	pisum sativ	958	6	1.2	862	1	SLA2_BACAN	P94217	bacillus an
886	6	1.2	723	1	PAL1_PEA	P04593	pisum sativ	959	6	1.2	867	1	POL_MPMV	P07572	simian maso
887	6	1.2	725	1	PAL2_PEA	P27990	medicago sa	960	6	1.2	867	1	POL_SRYV	P04025	simian retr
888	6	1.2	725	1	PALY_MEDSA	P45714	trifolium s	961	6	1.2	867	1	POL_SRYV2	P15157	simian retr
889	6	1.2	727	1	IF2M_BOVIN	P46198	bos taurus	962	6	1.2	868	1	SRYC_DROME	P15619	drosophila
890	6	1.2	728	1	CATB_ASPFU	Q92405	aspergillus	963	6	1.2	870	1	BCAI_HUMAN	P56945	homo sapien
891	6	1.2	729	1	YP65_CABEL	Q09214	caenorhabdi	964	6	1.2	870	1	COYG_CABEL	Q22498	caenorhabdi
892	6	1.2	730	1	CATR_ASPNG	P55314	aspergillus	965	6	1.2	870	1	DYN2_HUMAN	P05070	homo sapien
893	6	1.2	735	1	DBH4_MOUSE	P51660	mus musculu	966	6	1.2	870	1	DYN2_MOUSE	P39054	mus musculu
894	6	1.2	738	1	YO13_BP12	P42548	bacterioph	967	6	1.2	870	1	DYN2_RAT	P39054	rattus norv
895	6	1.2	742	1	DHET_ACENC	P18278	acetobacter	968	6	1.2	870	1	POL_JSRV	P31623	sheep pulmo
896	6	1.2	743	1	KPC2_APLCA	Q16975	aplysia cal	969	6	1.2	874	1	BCAI_MOUSE	Q61140	mus musculu
897	6	1.2	743	1	VN02_CABEL	Q03606	caenorhabdi	970	6	1.2	878	1	SECA_ANTSP	Q04641	antihelmin
898	6	1.2	745	1	VAT_THEAC	Q05209	thermoplasm	971	6	1.2	878	1	SYL_TREDA	Q83595	treponema p
899	6	1.2	752	1	GC20_YEAST	P43525	saccharomyc	972	6	1.2	887	1	MCW2_DROME	P49755	drosophila
900	6	1.2	752	1	XDHA_ECOS7	Q8r6c7	escherichia	973	6	1.2	888	1	POL_SMYVH	P03364	squirtrel mo
901	6	1.2	752	1	XDHA_ECOLI	Q46799	escherichia	974	6	1.2	892	1	YLB3_CABEL	P46378	caenorhabdi
902	6	1.2	758	1	GCR_ONCWY	P48843	oncothychnu	975	6	1.2	895	1	SECA_CVACA	Q19911	cyanidium c
903	6	1.2	763	1	TSHR_BOVIN	Q27987	bos taurus	976	6	1.2	897	1	CHIA_ECOLI	P13656	escherichia
904	6	1.2	764	1	TSHR_CANPA	P14763	canis famil	977	6	1.2	899	1	POL_MMYVB	P03365	mouse mamma
905	6	1.2	764	1	TSHR_MOUSE	P14750	mus musculu	978	6	1.2	901	1	POD2_TREPA	O81378	creponema p
906	6	1.2	764	1	TSHR_RAT	P21463	rattus norv	979	6	1.2	902	1	NRC4_HUMAN	O14974	homo sapien
907	6	1.2	764	1	TSHR_SHEEP	P56495	ovis aries	980	6	1.2	903	1	YB36_METUA	Q38556	metanococc
908	6	1.2	765	1	TREY_MYCTU	Q10768	haemophilus	981	6	1.2	905	1	HXAI_HAEIN	P44602	haemophilus
909	6	1.2	766	1	BCSB_SALTY	Q08290	salmonella	982	6	1.2	909	1	RP52_ARATH	P42484	arabidopsis

983	6	1.2	909	1	SBCC DEIRA	Q9T44 deiococcus
984	6	1.2	915	1	YLS4 CASEL	P4389 caenorhabdi
985	6	1.2	917	1	IL6B MOUSE	O00560 mus musculu
986	6	1.2	918	1	KRCM MOUSE	O62101 mus musculu
987	6	1.2	919	1	ATC1_HUMAN	P88199 homo sapien
988	6	1.2	919	1	ATC1_RAT	O64566 ratu
989	6	1.2	919	1	SYNP_HUMAN	Q9Y618 homo sapien
990	6	1.2	921	1	CA19_HUMAN	P08449 homo sapien
991	6	1.2	923	1	SECA_CAUCR	P38380 caulobacter
992	6	1.2	929	1	BLJ3 MOUSE	O9W992 mus musculu
993	6	1.2	933	1	PLD2_MOUSE	P97813 mus musculu
994	6	1.2	933	1	PLD2_RAT	P70496 ratu
995	6	1.2	934	1	CLTC_RAT	P27653 r c-1-tera
996	6	1.2	940	1	SEAL_MYCLE	P57936 mycobacteri
997	6	1.2	943	1	LBPA_NEIMB	O06379 neisseria m
998	6	1.2	944	1	LBPA_NEIMA	O93K4 neisseria m
999	6	1.2	948	1	SECA_MYCBO	P1494 mycobacteri
1000	6	1.2	949	1	SEAL_MYCTU	O05885 mycobacteri

ALIGNMENTS

RESULT 1

ANIA_NEIGO	STANDARD;	PRT;	392 AA.
ID ANIA_NEIGO	002219;		
DT 01-JUN-1994 (Rel. 29, Created)			
DT 01-JUN-1994 (Rel. 29, Last sequence update)			
DT 15-SEP-2003 (Rel. 42, Last annotation update)			
DE Major outer membrane protein Pan 1 precursor.			
GN ANIA.			
OS Neisseria gonorrhoeae.			
OC Bacteria: Proteobacteria; Betaproteobacteria; Neisseriales;			
OC Neisseriaceae; Neisseria.			
OK NCBI_TaxID=485;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=RI0;			
RX MEDLINE=93014187; PubMed=1383156;			
RA Hoehn G.T., Clark V.L.;			
RT "Isolation and nucleotide sequence of the gene (ania) encoding the			
RT major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae.";			
RL Infect. Immun. 60:4695-4703 (1992).			
CC [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=ATCC 33084 / F62;			
RX MEDLINE=93014188; PubMed=1398981;			
RA Hoehn G.T., Clark V.L.;			
RT "The major anaerobically induced outer membrane protein of Neisseria			
RT gonorrhoeae, Pan 1, is a lipoprotein.";			
RL Infect. Immun. 60:4704-4708 (1992).			
CC [1]			
CC SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid			
CC anchor (probable).			
CC [1]			
CC INDUCTION: BY ANAEROBICITY.			
CC [1]			
CC SIMILARITY: Contains 2 plastocyanin-like domains.			
CC [1]			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/			
CC or send an email to license@ebi.ac.uk).			
CC [1]			
CC EMBL: M97926; AAA25462.1; -			
DR PIR: A49208; A49208.			
DR PDB: 1KBV; 27-FEB-02.			
DR PDB: 1KBW; 27-FEB-02.			
DR InterPro: IPR001117; Cu-oxidase.			
DR InterPro: IPR001287; CUNO2_reductase.			
DR Pfam: PF00394; Cu-oxidase; 2.			

DR PRINTS; PR00695; CUNO2RDYASE.

DR POSITIVE; PS00013; PROKAR LIPOPROTEIN; 1.

KM Repeat; Outer membrane; Lipoprotein; Signal; 3D-structure.

FT	CHAIN	1	18	MAJOR OUTER MEMBRANE PROTEIN PAN 1.
FT	CHAIN	1	18	
FT	LIPID	19	392	
FT	DOMAIN	101	195	N-ACYL DIGLYCERIDE (PROBABLE).
FT	DOMAIN	245	346	PLASTOCYANIN-LIKE 1.
FT	DOMAIN	368	387	PLASTOCYANIN-LIKE 2.
FT	REPEAT	368	372	4 X 5 AA TANDEM REPEATS OF A-A-S-A-P.
FT	REPEAT	373	377	
FT	REPEAT	378	382	
FT	REPEAT	383	387	
FT	REPEAT	387	392	

SO SEQUENCE 392 AA; 40954 MW; A4707CC878923C97 CRC64;

Query Match 6.8%; Score 34; DB 1; Length 392;

Best Local Similarity 100.0%; Pred.No.5.5e-27;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 178 YHCAPVPGVHIANGYGLVPEKGLPKVDKE 211

Db 173 YHCAPVPGVHIANGYGLVPEKGLPKVDKE 206

RESULT 2

NIR_ALCXX	STANDARD;	PRT;	330 AA.
ID NIR_ALCXX	001445;		
DT 28-FEB-2003 (Rel. 41, Created)			
DT 28-FEB-2003 (Rel. 41, Last sequence update)			
DT 15-SEP-2003 (Rel. 42, Last annotation update)			
DE Copper-containing nitrite reductase (EC 1.7.2.1) (Cu-NIR).			
GN NIRK.			
OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).			
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC Alcaligenaceae; Achromobacter.			
OK NCBI_TaxID=515;			
RN [1]			
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).			
RC STRAIN=NCIMB 11015.			
RA Dodd F.E., Hasnain S.S., Abraham Z.H.L., Eady R.R., Smith B.E.;			
RT "Structures of a blue-copper nitrite reductase and its substrate-bound			
RT complex.";			
RL Acta Crystallogr. D 53:406-418 (1997).			
CC [2]			
CC CHARACTERIZATION.			
RC STRAIN=NCIMB 11015.			
RX MEDLINE=94059002; PubMed=8240262;			
RA Abraham Z.H.L., Lowe D.J., Smith B.E.;			
RT "Purification and characterization of the dissimilatory nitrite			
RT reductase from Alcaligenes xylosoxydans subsp. xylosoxydans			
RT (N.C.I.M.B. 11015): evidence for the presence of both type 1 and type			
RT 2 copper centres.";			
RL Biochem. J. 295:587-593 (1993).			
CC [1]			
CC CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferricytochrome c =			
CC nitrite + ferrocyclochrome c.			
CC [1]			
CC COFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE			
CC II COPPER. PAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER.			
CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS			
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.			
CC PSEUDOAQUINONE IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN			
CC VITRO.			
CC [1]			
CC PATHWAY: Nitrate assimilation (denitrification).			
CC [1]			
CC SUBUNIT: Homotrimer.			
CC [1]			
CC SUBCELLULAR LOCATION: Periplasmic (By similarity).			
CC [1]			
CC DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR			
CC ELECTRON TRANSFER FROM PSEUDOAQUINONE TO THE TYPE II COPPER SITE OF			
CC NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE REDUCTION			
CC OF NITRITE.			
CC [1]			
CC SIMILARITY: Contains 2 plastocyanin-like domains.			
CC PDB: INDR; 07-JUL-97.			
DR PDB: INDS; 07-JUL-97.			
DR InterPro: IPR001117; Cu-oxidase.			


```
DR InterPro: IPR001287; CuNO2_reductase.
DR Pfam: PF00394; Cu-oxidase; 2.
DR PRINTS: PR00695; CONO2RDTASE.
KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
KW Nitrate assimilation; Repeat; Periplasmic; 3d-structure.
FT DOMAIN 1 165 PLASTOCYANIN-LIKE 1.
FT METAL 166 330 PLASTOCYANIN-LIKE 2.
FT METAL 85 85 COPPER (TYPE 1).
FT METAL 90 90 COPPER (TYPE 2).
FT METAL 125 125 COPPER (TYPE 2).
FT METAL 126 126 COPPER (TYPE 1).
FT METAL 135 135 COPPER (TYPE 1).
FT METAL 140 140 COPPER (TYPE 1).
FT METAL 296 296 COPPER (TYPE 2).
SQ SEQUENCE 330 AA; 34413 MW; E2C38C3A2CEBFCE8 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0003; Length 330;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDPHATG 148
DB 83 MPNVDPHATG 94

RESULT 3
NIR_PSECL STANDARD; PRT; 363 AA.
AC 006006;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
GN NIRK.
OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=333;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.
RC STRAIN=ATCC 13985;
RX MEDLINE=93356602; PubMed=8352648;
RA Glockner A.B., Juengst A., Zumft W.G.;
RT "Copper-containing nitrite reductase from Pseudomonas aureofaciens is functional in a mutationally cytochrome cdl-free background (Nirs-)"
RT of Pseudomonas stutzeri.";
RL Arch. Microbiol. 160:18-26(1993).
CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferri-cytochrome c =
CC nitrite + ferri-cytochrome c.
CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
CC II COPPER; FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
CC PSEUDOMAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
CC VITRO.
CC -1- PATHWAY: Nitrate assimilation (denitrification).
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
CC ELECTRON TRANSFER FROM PSEUDOMAZURIN TO THE TYPE II COPPER SITE
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
CC REDUCTION OF NITRITE.
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z21945; CAA79939.1; -
```

```
DR PIR; S32112; S32112.
DR HSSP; P38501; 1A58.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR001287; CuNO2_reductase.
DR Pfam: PF00394; Cu-oxidase; 2.
DR PRINTS: PR00695; CONO2RDTASE.
KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
KW Nitrate assimilation; Repeat; Periplasmic; Signal.
FT SIGNAL 1 24
FT CHAIN 25 363 COPPER-CONTAINING NITRITE REDUCTASE.
FT DOMAIN 25 193 PLASTOCYANIN-LIKE 1.
FT METAL 194 363 PLASTOCYANIN-LIKE 2.
FT METAL 113 113 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 118 118 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 153 153 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 154 154 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 163 163 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 168 168 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 324 324 COPPER (TYPE 2) (BY SIMILARITY).
SQ SEQUENCE 363 AA; 39248 MW; 50DDB60CC4DC3E00 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00032; Length 363;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDPHATG 148
DB 111 MPNVDPHATG 122

RESULT 4
NIR_RHIHE STANDARD; PRT; 377 AA.
AC 060214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
GN NIRK.
OS Rhizobium hedyarai.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=50338;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCNT1;
RX MEDLINE=97055705; PubMed=8699992;
RA Toftanin A., Wu Q., Maskus M., Casella S., Abruna H.D.,
RA Shapleigh J.P.;
RT "Characterization of the gene encoding nitrite reductase and the
RT physiological consequences of its expression in the nondenitrifying
RT Rhizobium hedyarai strain HCNT1.";
RL Appl. Environ. Microbiol. 62:4019-4025(1996).
CC -1- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferri-cytochrome c =
CC nitrite + ferri-cytochrome c.
CC -1- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
CC II COPPER; FAD. TYPE I COPPER IS BOUND WITHIN A SINGLE MONOMER,
CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
CC OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
CC PSEUDOMAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
CC VITRO (BY SIMILARITY).
CC -1- PATHWAY: Nitrate assimilation (denitrification).
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -1- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
CC ELECTRON TRANSFER FROM PSEUDOMAZURIN TO THE TYPE II COPPER SITE
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
CC REDUCTION OF NITRITE.
CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U65658; AAB05880.1; -.
 DR HSP; P38501; 1AS7.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR001287; CuNO2_reductase.
 DR InterPro; IPR006311; Tat.
 DR Pfam; PF00394; Cu-oxidase; 2.
 DR PRINTS; PR00695; CUNO2RDTASE.
 DR TIGR/MS; TIGR01409; Tat_signal_seq; 1.
 KM Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
 KM Nitrate assimilation; Repeat; Periplasmic; Signal.
 FT SIGNAL 1 46
 FT CHAIN 1 46 POTENTIAL.
 FT DOMAIN 47 377 COPPER-CONTAINING NITRITE REDUCTASE.
 FT DOMAIN 99 194 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 259 360 PLASTOCYANIN-LIKE 2.
 FT METAL 132 132 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 137 137 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 172 172 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 173 173 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 182 182 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 187 187 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 343 343 COPPER (TYPE 2) (BY SIMILARITY).
 SQ SEQUENCE 377 AA; 40719 MW; C7AFEE9AE639E2E8 CRC64;

Query Match 2.0%; Score 10; DB 1; Length 377;
 Best Local Similarity 100.0%; Pred. No. 0.039;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 252 VGALTGENAL 261
 Db 261 VGALTGENAL 270

RESULT 5
 RCSF_ECOLI
 ID RCSF_ECOLI STANDARD; PRT; 134 AA.
 AC P28633;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein rcsf.
 GN RCSF OR B0196 OR C0237.
 OS Escherichia coli, and
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=93094132; PubMed=1459951;
 RA Gervais F.G., Drapeau G.R.;
 RT "Identification, cloning, and characterization of rcsf, a new
 RT regulator gene for exopolysaccharide synthesis that suppresses the
 RT division mutation ftsZ84 in Escherichia coli K-12.";
 RL J. Bacteriol. 174:8016-8022(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA Miyamoto K.;
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Coliado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
 RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
 RA Mizuno T., Makino K., Nakata A., Yura T., Samped G., Mizobuchi K.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of the
 RT 4.0-6.0 min (189,987-281,416bp) region.";
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
 RA Davis K., Federspiel N., Hyman R., Kallman S., Komp C., Kundi O.,
 RA Laskhari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=06:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=2388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Raske D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blatter F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC - FUNCTION: Stimulates colanic acid capsule synthesis. May
 CC promote the phosphorylation of rcsB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; L04474; AAA24508.1; -.
 DR EMBL; D15061; BAA03656.1; -.
 DR EMBL; AEO00128; AAC73307.1; -.
 DR EMBL; D83536; BAA77873.1; -.
 DR EMBL; U70214; AAB08624.1; -.
 DR EMBL; AEO16755; AAN78729.1; -.
 DR PIR; D64744; D64744.
 DR EcoGene; Egl1502; rcsf.
 KM Bacterial capsule; Complete proteome.
 FT CONFLICT 37 38 EP -> DA (IN REF. 1 AND 4).
 FT CONFLICT 46 46 MISSING (IN REF. 1 AND 4).
 SQ SEQUENCE 134 AA; 14163 MW; 824F82151C07BB41 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 134;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 17 ALMLSGCS 24
 Db 10 ALMLSGCS 17

RESULT 6
 PORA_METUA
 ID PORA_METUA STANDARD; PRT; 389 AA.
 AC O57715;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pyruvate synthase subunit porA (EC 1.2.7.1) (Pyruvate oxidoreductase
 DE alpha chain) (POR) (Pyruvic-ferredoxin oxidoreductase alpha subunit).
 GN PORA OR MJ0267.
 OS Methanococcus jannaschii.

```

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
  [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JUL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Bull C.U., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sultun G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervilavag A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Ruhnman J.L., Nguyen D.,
RA Uetereback I.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
  jamaeschi";
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: Pyruvate + CoA + oxidized ferredoxin = acetyl-
CC -1- SUBUNIT: HETEROTETRAMER OF ONE ALPHA, ONE BETA, ONE DELTA AND ONE
CC GAMMA CHAIN (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67482; AAB98254.1; -.
DR PIR; D64333; D64333.
DR TIGR; MJO267; -.
DR InterPro; IPR002880; PRO_N.
DR Pfam; PF01855; PRO_N; 1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 389 AA; 42971 MW; 2CAA3DDE3EB0F5A9 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 389;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 FNKGALGI 347
  |||||
Db 332 FNKGALGI 339

RESULT 7
TIG-THETN STANDARD; PRT; 432 AA.
AC ORFC26;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Trigger factor (TF).
DE TIG OR TTB0624.
OS Thermoaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoaerobacteriales;
OC Thermoaerobacteriaceae; Thermoaerobacter.
OX NCBI_TaxID=119072;
  [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -1- FUNCTION: Involved in protein export. Acts as a chaperone by
CC maintaining the newly synthesized protein in an open conformation
CC (By similarity).

```

```

CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PIPIASE FAMILY. TIG SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB013031; AM23893.1; -.
DR HAMAP; MF_00303; -.
DR InterPro; IPR001179; FKBP_PIPIASE.
DR InterPro; IPR005215; TIG_fac.
DR TIGRPFAM; TIGR00115; tlg_1.
DR PROSITE; PS00453; FKBP_PIPIASE_1; FALSE_NEG.
DR PROSITE; PS00454; FKBP_PIPIASE_2; FALSE_NEG.
DR PROSITE; PS00059; FKBP_PIPIASE_3; 1.
KW Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.
FT DOMAIN 163 248 PIPIASE, FKBP-TYPE.
SQ SEQUENCE 432 AA; 49762 MW; F7DC7DC821072E15 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 432;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 PEGKGEN 303
  |||||
Db 179 PEGKGEN 186

RESULT 8
SAP-CHICK
ID -SAP-CHICK STANDARD; PRT; 518 AA.
AC O13035;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Proactivator polypeptide precursor [Contains: Saposin A; Saposin B;
DE Saposin C; Saposin D].
GN PSAP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
  [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 194-203.
RC TISSUE=Brain, and Liver;
RX MEDLINE=98129745; PubMed=9461526;
RA Azuma N., Seo H.-C., Lie O., Fu Q., Gould R.M., Hiraiwa M., Burt D.W.,
RA Paton I.R., Morrice D.R., O'Brien J.S., Kishimoto Y.;
RT "Cloning, expression and map assignment of chicken prosaposin.";
RL Biochem. J. 330:321-327(1998).
  [2]
RP SEQUENCE FROM N.A.
RA Altman N., Horowitz M.;
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THE LYSOSOMAL DEGRADATION OF SPHINGOLIPIDS TAKES PLACE
CC BY THE SEQUENTIAL ACTION OF SPECIFIC HYDROLASES. SOME OF THESE
CC ENZYMES REQUIRE SPECIFIC LOW-MOLECULAR MASS, NON-ENZYMIC PROTEINS:
CC THE SPHINGOLIPIDS ACTIVATOR PROTEINS (COPROTEINS) (BY SIMILARITY).
CC -1- FUNCTION: SAPOIN A AND SAPOIN C STIMULATE THE HYDROLYSIS OF
CC GLUCOSYLCERAMIDE BY BETA-GLUCOSYLCERAMIDASE (EC 3.2.1.45) AND
CC GALACTOSYLCERAMIDE BY BETA-GALACTOSYLCERAMIDASE (EC 3.2.1.46).
CC SAPOIN-C APPARENTLY ACTS BY COMBINING WITH THE ENZYME AND ACIDIC
CC LIPID TO FORM AN ACTIVATED COMPLEX, RATHER THAN BY SOLUBILIZING
CC THE SUBSTRATE (BY SIMILARITY).
CC -1- FUNCTION: SAPOIN B STIMULATES THE HYDROLYSIS OF GALACTO-
CC CEREBROSIDE SULFATE BY ARYL SULFATASE A (EC 3.1.6.8), GMI
CC GANGLIOSIDES BY BETA-GALACTOSIDASE (EC 3.2.1.23) AND
CC GLOBOTRIASOLYCERAMIDE BY ALPHA-GALACTOSIDASE A (EC 3.2.1.22).
CC SAPOIN-B FORMS A SOLUBILIZING COMPLEX WITH THE SUBSTRATES OF THE

```

```

CC SPHINGOLIPID HYDROLASES (BY SIMILARITY).
CC -1- FUNCTION: SAPOSIN D IS A SPECIFIC SPHINGOMYELIN PHOSPHODIESTERASE
CC ACTIVATOR (EC 3.1.4.12) (BY SIMILARITY).
CC -1- SUBUNIT: SAPOSIN B IS A HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Lysosomal (By similarity).
CC -1- PFM: THIS PRECURSOR IS PROTEOLYTICALLY PROCESSED TO A SMALL
CC PEPTIDES, WHICH ARE SIMILAR TO EACH OTHER AND ARE SPHINGOLIPID
CC HYDROLASE ACTIVATOR PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: Contains 4 saposin A-type domains.
CC -1- SIMILARITY: Contains 4 saposin B-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB003471; BAA19914.1; -.
CC EMBL; AF108656; AAF05899.1; -.
CC InterPro; IPR003119; SApA.
CC InterPro; IPR000004; SApB.
CC DR Pfam; PF02199; SApA; 2.
CC DR Pfam; PF05184; SApB.1; 4.
CC DR Pfam; PF03489; SApB.2; 4.
CC DR ProDom; PD001732; SApB sub.3.
CC KW Signal; Glycoprotein; Lysosome; Sphingolipid metabolism; Repeat;
CC Gln2-gangliosidosis.
CC FT SIGNAL 1 17 POTENTIAL.
CC FT PROPEP 18 60
CC FT CHAIN 61 143 SAPOSIN A.
CC FT PROPEP 145 193
CC FT CHAIN 194 276 SAPOSIN B.
CC FT PROPEP 278 305
CC FT CHAIN 307 387 SAPOSIN C.
CC FT PROPEP 389 398
CC FT CHAIN 399 480
CC FT PROPEP 482 518 SAPOSIN D.
CC FT DOMAIN 22 55
CC FT DOMAIN 60 143 SAPOSIN-LIKE TYPE A 1.
CC FT DOMAIN 193 276 SAPOSIN-LIKE TYPE B 1.
CC FT DOMAIN 307 388 SAPOSIN-LIKE TYPE B 2.
CC FT DOMAIN 399 480 SAPOSIN-LIKE TYPE B 3.
CC FT DOMAIN 485 518 SAPOSIN-LIKE TYPE A 2.
CC FT DISULFID 64 139 BY SIMILARITY.
CC FT DISULFID 67 133 BY SIMILARITY.
CC FT DISULFID 95 107 BY SIMILARITY.
CC FT DISULFID 197 273 BY SIMILARITY.
CC FT DISULFID 200 267 BY SIMILARITY.
CC FT DISULFID 229 240 BY SIMILARITY.
CC FT DISULFID 311 384 BY SIMILARITY.
CC FT DISULFID 314 378 BY SIMILARITY.
CC FT DISULFID 342 353 BY SIMILARITY.
CC FT DISULFID 403 476 BY SIMILARITY.
CC FT DISULFID 406 470 BY SIMILARITY.
CC FT DISULFID 434 445 BY SIMILARITY.
CC FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 94 94 R -> T (IN REF. 2).
CC FT CONFLICT 486 486 E -> D (IN REF. 2).
CC SEQUENCE 518 AA; 57601 MW; B803000E891C3963 CRC64;

```

Query Match 1.6%; Score 8; DB 1; Length 518;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 VKMETVEK 93
 |||||
 Db 293 VKMETVEK 300

```

RESULT 9
ID yf1c_BACSU STANDARD; PRT; 604 AA.
AC P54719;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein yf1c.
GN yf1c.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96262713; PubMed=8704981;
RA Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.,
RT "Determination of a 12 kb nucleotide sequence around the 76 degrees
RL Microbiology 142:1417-1421 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Borries R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brueschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dueterhoff A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Hage K., Hatech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holseppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Katsuhara Y., Klaerr-Bianchini M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestel D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Satale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serot S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzsaecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D50543; BAA09107.1; -.
CC EMBL; Z99108; CAB12651.1; -.
CC DR PIR; P69802; P69802.
CC DR HSSP; P13569; INBD.
CC DR Subtilin; BG1850; yf1c.
CC InterPro; IPR003593; AAA_ATPase.

```

```
DR InterPro: IPR001140; ABC_TM_transpt.  
DR InterPro: IPR003439; ABC_transporter.  
DR Pfam: PF00664; ABC_membrane.1.  
DR Pfam: PF00005; ABC_tran.1.  
DR ProDom: PD000006; ABC_transporter.1.  
DR SMART: SM00382; AAA; T.  
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.  
DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.  
KW Hypothetical protein; ATP-binding; Transport; Transmembrane;  
KW Complete proteome.  
FT TRANSMEM 50 70 POTENTIAL.  
FT TRANSMEM 86 106 POTENTIAL.  
FT TRANSMEM 172 192 POTENTIAL.  
FT TRANSMEM 288 308 POTENTIAL.  
FT TRANSMEM 510 530 POTENTIAL.  
FT NP_BIND 399 406 ATP (POTENTIAL).  
SQ SEQUENCE 604 AA; 67396 MW; 1998B4669A2E6133 CRC64;  
  
Query Match 1.4%; Score 8; DB 1; Length 604;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 40 AAKTANA 47  
DB 474 AAKTANA 481  
  
RESULT 10  
C555_CHLIT STANDARD; PRT; 86 AA.  
ID C555_CHLIT STANDARD; PRT; 86 AA.  
AC P00123;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Cytochrome c-555 (C555).  
OS Chlorobium limicola f.sp. thiosulfatophilum.  
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
OC Chlorobium.  
OX NCBI_TaxId=115852;  
RN [1]  
RP SEQUENCE:  
RX MEDLINE=77087088; PubMed=188412;  
RA van Beekman J., Ambler R.P., Meyer T.E., Kamen M.D., Olson J.M.,  
RA Shaw E.K.;  
RT "The amino acid sequences of the cytochromes c-555 from two green  
RT sulphur bacteria of the genus Chlorobium.";  
RL Biochem. J. 159:757-774(1976).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RA MEDLINE=78094383; PubMed=202947;  
RA Korszun Z.R., Saleme F.R.;  
RT "Structure of cytochrome c555 of Chlorobium thiosulfatophilum:  
RT primitive low-potential cytochrome c.";  
RL Proc. Natl. Acad. Sci. U.S.A. 74:5244-5247(1977).  
CC -1- FUNCTION: This basic c-type monoheme cytochrome has been found  
CC exclusively in the green photosynthetic bacteria, although its  
CC role in bacterial photosynthesis is not established. It has an  
CC unusually low redox potential compared with mitochondrial  
CC cytochrome c. It is reactive with cytochrome c oxidases but not  
CC with reductases.  
DR HSP; P11732; 1CC5.  
DR InterPro: IPR003088; Cyt_C1.  
DR InterPro: IPR000345; CytC_heme_bind.  
DR Pfam: PF00034; Cytochrome_c; 1.  
DR PROSITE: PS00190; CYTOCHROME_C; 1.  
KW Electron transport; Photosynthesis; Heme.  
FT ELECTRON_TRANSPORT; PHOTOSYNTHESIS; HEME.  
FT BINDING 14 14 HEME (COVALENT).  
FT BINDING 17 17 HEME (COVALENT).  
FT METAL 18 18 IRON (HEME AXIAL LIGAND).  
FT METAL 60 60 IRON (HEME AXIAL LIGAND).  
SQ SEQUENCE 86 AA; 8780 MW; 0882D21350DB9D4E CRC64;  
  
Query Match 1.4%; Score 7; DB 1; Length 86;
```

```
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 397 AGKATYD 403  
DB 5 AGKATYD 11  
  
RESULT 11  
C555_CHLITE STANDARD; PRT; 108 AA.  
ID C555_CHLITE STANDARD; PRT; 108 AA.  
AC Q8KG93;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cytochrome c-555 precursor (C555).  
GN CT0075.  
OS Chlorobium tepidum.  
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
OC Chlorobium.  
OX NCBI_TaxId=1097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=7LS / ATCC 49652 / DSM 12025;  
RX MEDLINE=22103685; PubMed=12093901;  
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
RA Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,  
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Raddue D.,  
RA Vamathevan J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
RT "The complete genome sequence of Chlorobium tepidum 7LS, a  
RT photosynthetic, anaerobic, green-sulfur bacterium.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
CC -1- FUNCTION: This basic c-type monoheme cytochrome has been found  
CC exclusively in the green photosynthetic bacteria, although its  
CC role in bacterial photosynthesis is not established. It has an  
CC unusually low redox potential compared with mitochondrial  
CC cytochrome c. It is reactive with cytochrome c oxidases but not  
CC with reductases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: AE012787; AAM71323.1; -  
DR TIGR: CT0075; -  
DR InterPro: IPR003088; Cyt_C1.  
DR InterPro: IPR002323; Cyt_C1E.  
DR InterPro: IPR000345; CytC_heme_bind.  
DR Pfam: PF00034; Cytochrome_c; 1.  
DR PRINTS: PR00607; CYTOCHROME_C.  
DR PROSITE: PS00190; CYTOCHROME_C; 1.  
KW Electron transport; Photosynthesis; Heme; Signal; Complete proteome.  
FT SIGNAL 1 22 BY SIMILARITY.  
FT CHAIN 23 108 CYTOCHROME_C-555.  
FT BINDING 36 36 HEME (COVALENT) (BY SIMILARITY).  
FT BINDING 39 39 HEME (COVALENT) (BY SIMILARITY).  
FT METAL 40 40 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 82 82 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 108 AA; 10892 MW; 9711CB0A43B8825C CRC64;  
  
Query Match 1.4%; Score 7; DB 1; Length 108;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 397 AGKATYD 403  
DB 5 AGKATYD 11
```

Db 27 AGKATYD 33

RESULT 12

RBS_PROHO

ID RBS_PROHO STANDARD; PRT; 109 AA.

AC P27569;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Ribulose biphosphate carboxylase small chain (EC 4.1.1.39) (RUBISCO small subunit).

GN CBBS OR RBGS.

OS Prochlorothrix hollandica.

OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorothrixaceae;

OC Prochlorothrix.

OC NCBI_TaxID=1223;

OK NCBI_TaxID=1223;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91251137; PubMed=1904095;

RA Morden C.W., Golden S.S.;

RT "Sequence analysis and phylogenetic reconstruction of the gene encoding the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase from the chlorophyll b-containing prokaryote *Prochlorothrix hollandica*."

RL J. Mol. Evol. 33:379-395(1991).

CC -1- FUNCTION: RUBISCO catalyzes two reactions: the carboxylation of D-ribulose 1,5-bisphosphate, the primary event in photosynthetic carbon dioxide fixation, as well as the oxidative fragmentation of the pentose substrate in the photorespiration process. Both reactions occur simultaneously and in competition at the same active site.

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-phospho-D-glycerate.

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) = 3-phospho-D-glycerate + 2-phosphoglycolate.

CC -1- SUBUNIT: 8 large chains + 8 small chains.

CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC DR EMBL; X57359; CAA0633.1; -.

CC DR PIR; S16437.

CC DR InterPro; IPR000894; RUBISCO_small.

CC DR Pfam; PF00101; RUBISCO_small.1.

CC DR PRINTS; PRO0152; RUBISCO_SMALL.

CC DR ProDom; PD000290; RUBISCO_small.1.

CC KM Photosynthesis; Carbon dioxide fixation; Photorespiration; Lysase; Oxidoreductase; Monooxygenase.

CC SQ SEQUENCE 109 AA; 12963 MW; 9A17AE63ACDD5BE CRC64;

Query Match 1.4%; Score 7; DB 1; Length 109;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 LSDQOIA 471

DB 19 LSDQOIA 25

RESULT 13

RBS_SYNP2

ID RBS_SYNP2 STANDARD; PRT; 111 AA.

AC Q44178;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Ribulose biphosphate carboxylase small chain (EC 4.1.1.39) (RUBISCO small subunit).

GN CBBS OR RBGS.

OS *Synechococcus* sp. (strain PCC 7002) (*Agmenellum quadruplicatum*).

OC Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.

OC NCBI_TaxID=32049;

OK NCBI_TaxID=32049;

RN [1]

RP SEQUENCE FROM N.A.

RA Akiyama H., Kanai S., Hirano M., Sugimoto M., Kiyohara M.;

RA Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: RUBISCO catalyzes two reactions: the carboxylation of D-ribulose 1,5-bisphosphate, the primary event in photosynthetic carbon dioxide fixation, as well as the oxidative fragmentation of the pentose substrate in the photorespiration process. Both reactions occur simultaneously and in competition at the same active site.

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-phospho-D-glycerate.

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) = 3-phospho-D-glycerate + 2-phosphoglycolate.

CC -1- SUBUNIT: 8 large chains + 8 small chains.

CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC DR EMBL; D13971; BAA03078.1; -.

CC DR InterPro; IPR000894; RUBISCO_small.

CC DR Pfam; PF00101; RUBISCO_small.1.

CC DR PRINTS; PRO0152; RUBISCO_SMALL.

CC DR ProDom; PD000290; RUBISCO_small.1.

CC KM Photosynthesis; Carbon dioxide fixation; Photorespiration; Lysase; Oxidoreductase; Monooxygenase.

CC SQ SEQUENCE 111 AA; 13212 MW; 182FA0950AC8EF96 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 111;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 LSDQOIA 471

DB 19 LSDQOIA 25

RESULT 14

YC39_AQUAE

ID YC39_AQUAE STANDARD; PRT; 111 AA.

AC 067264;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein AQ_1239.

GN AQ_1239.

OS Aquifex aeolicus.

OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.

OC NCBI_TaxID=63363;

OK NCBI_TaxID=63363;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VF5;

RC MEDLINE=98196666; PubMed=9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., RA Graham D.E., Overbeek R., Shead M.A., Keller M., Anay M., Huber R., RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;

RT "The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*."

RT Nature 392:353-358(1998).

RL

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000730; AAC07249.1; -
DR PIR: A70407; A70407.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 111 AA; 13020 MW; 262042D32BD8A3A2 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 228 OGLOPFD 234
DB 28 OGLOPFD 34

RESULT 15
RS9_LACIA STANDARD; PRT; 130 AA.
AC 09CDE7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S9.
GN RPS1 OR L12253.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wnucker P., Mager S., Jallion O., Malarne K.,
RA Weisenbach J., Erlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis esp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -1- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE006454; AAK06351.1; -
DR PIR: E86906; E86906.
DR HAMAP: MF_00532; -; 1.
DR InterPro: IPR000754; Ribosomal_S9.
DR Pfam: PF00380; Ribosomal_S9; 1.
DR ProDom: PD001627; Ribosomal_S9; 1.
DR PROSITE: PS00360; RIBOSOMAL_S9; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 130 AA; 14098 MW; CAF9BC8DCD531BBB CRC64;

Query Match 1.4%; Score 7; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 447 GRITVG 453
DB 26 GRITVG 32

RESULT 16
GREX_BRUME
GREX_BRUME

ID GREX_BRUME STANDARD; PRT; 157 AA.
AC 08YID6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription elongation factor greA (Transcript cleavage factor
DE greA).
GN GREX OR BME10508 OR BR1504.
OS Brucella melitensis, and
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459, 29461;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B. melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=B. suis; STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Frazer C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC -1- FUNCTION: Necessary for efficient RNA polymerase transcription
CC elongation past template-encoded arresting sites. The arresting
CC sites in DNA have the property of trapping a certain fraction of
CC elongating RNA polymerases that pass through, resulting in locked
CC ternary complexes. Cleavage of the nascent transcript by cleavage
CC factors such as greA or greB allows the resumption of elongation
CC from the new 3' terminus. GreA releases sequences of 2 to 3
CC nucleotides (by similarity).
CC
CC -1- SIMILARITY: BELONGS TO THE GREX/GREB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE004493; AAL51689.1; -
DR PIR: AF3145; AF3145.
DR HAMAP: BR1504; -
DR TIGR: BR1504; -
DR InterPro: IPR006359; GreA.
DR Pfam: PF01272; GreA_Greb_1.
DR ProDom: PD004918; GreA_Greb_N; 1.
DR TIGRPFam: TIGR01462; greA; 1.
DR PROSITE: PS00829; GREAB_1; 1.
DR PROSITE: PS00830; GREAB_2; 1.
KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 46 74 COILED COIL (POTENTIAL).
SQ SEQUENCE 157 AA; 17528 MW; 5651BAF6D942B3F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 157;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 EGDTEV 127
 |||||
 Db 134 EGDTEV 140

RESULT 17

KCH_MOUSE STANDARD; PRT; 172 AA.
 ID KCH_MOUSE
 AC Q920E3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Voltage-gated potassium channel subfamily H member 5 (Ether-a-go-go
 DE potassium channel 2) (tag2) (fragment).
 GN KCNH5 OR EAG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Saganian M.J., Vega-Saenz de Miera B.C., Rudy B.;
 RT "Cloning of the mouse eag2 potassium channel".
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Pore-forming (alpha) subunit of voltage-gated potassium
 CC channel. Elicits a non-inactivating outward rectifying current
 CC (by similarity). Channel properties may be modulated by cAMP and
 CC subunit assembly.
 CC -1- SUBUNIT: The potassium channel is probably composed of a homo- or
 CC heterotrimeric complex of pore-forming alpha subunits that can
 CC associate with modulating beta subunits. Heteromultimer with
 CC KCNH1/EAG (by similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to the potassium channel family. H (Eag)
 CC subfamily.
 CC -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF309565; AAI09442.1; -
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR000700; PAS-assoC.
 DR InterPro: IPR00014; PAS_domain.
 DR Pfam: PF00785; PAC; 1.
 DR SMART: SM00086; PAC; 1.
 DR PROSITE: PS00113; PAC; 1.
 DR PROSITE: PS00112; PAS; PARTIAL.
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
 KW Potassium channel; Potassium; Potassium transport; Multigene family.
 FT NON TER 1 1
 FT DOMAIN <1 >172 CYTOPLASMIC (BY SIMILARITY).
 FT DOMAIN <1 59 PAS.
 FT DOMAIN 60 112 PAC.
 FT NON TER 172 172
 SQ SEQUENCE 172 AA; 20064 MM; 99B392D12917E67C CRC64;

Query Match 1.4%; Score 7; DB 1; Length 172;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 380 QEAPKTP 386

Db 166 QEAPKTP 172
 |||||

RESULT 18

EFA3_MOUSE STANDARD; PRT; 187 AA.
 ID EFA3_MOUSE
 AC O08545; O55217;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ephrin-A3 (EPH-related receptor tyrosine kinase ligand 3) (LERK-3)
 DE (EHK1 ligand) (EHK1-L) (fragment).
 GN EFN3 OR EPLG3 OR LERK3 OR EPL3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=98126446; PubMed=9465306;
 RA Cerretti D.P., Nelson N.;
 RT "Characterization of the genes for mouse LERK-3/Ephrin-A3 (Epl3),
 RT mouse LERK-4/Ephrin-A4 (Epl4), and human LERK-6/Ephrin-A2 (Eplg6):
 RT conservation of intron/exon structure.";
 RL Genomics 47:131-135 (1998).
 RN [2]
 RP SEQUENCE OF 17-78 FROM N.A.
 RX MEDLINE=97060319; PubMed=8903354;
 RA Flemken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;
 RT "Distinct and overlapping expression patterns of ligands for
 RT Eph-related receptor tyrosine kinases during mouse embryogenesis.";
 RL Dev. Biol. 179:382-401 (1996).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U92885; AAC39961.1; -
 DR EMBL: U90666; AAB50241.1; -
 DR WGD; MG1:106644; Etna3.
 DR InterPro: IPR001799; Ephrin.
 DR Pfam: PF00812; Ephrin; 1.
 DR PRINTS: PR01347; EPHRIN.
 DR PRODOM: PD002533; Ephrin; 1.
 DR PROSITE: PS01299; EPHRIN; 1.
 KW Glycoprotein; GPI-anchor.
 FT NON TER 1 1
 FT CARBOHYD 24 24 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 49 49 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 19 19 Y -> I (IN REF. 2).
 FT CONFLICT 33 34 EQ -> DR (IN REF. 2).
 FT CONFLICT 46 47 RT -> QP (IN REF. 2).
 FT CONFLICT 78 78 Y -> W (IN REF. 2).
 SQ SEQUENCE 187 AA; 21171 MM; CCE4915751760743 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 187;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 GPGGAE 154
 |||||
 Db 27 GPGGAE 33

RESULT 19
YEAST_ECOLI
ID YEAST_ECOLI STANDARD; PRT; 193 AA.
AC P76255;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical lipoprotein yeast precursor.
GN YEAST OR B1806 OR C2210.
OS Escherichia coli, and
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 21792;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Iseno K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizouchi K., Mori H., Mori T., Mocomu K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Sivasubraman S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horichi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RT DNA Res. 3:379-392(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raebou G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [4]
RP SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
RN [5]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation-
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000275; AAC74876.1; -
CC EMBL; D90824; BAA15610.1; -
CC EMBL; D90825; BAA15615.1; -
CC EMBL; AE016761; AAN80669.1; -
CC PIR; F64941; F64941.
CC Ecogen; E613511; Yeast.
CC InterPro; IPR004658; SLP.
CC Pfam; PF03843; SLP; 1.
CC TIGRfam; TIGR00752; slp; 1.
CC PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
CC Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
CC SIGNAL
FT CHAIN 1 22 POTENTIAL.
FT 23 193 HYPOTHETICAL LIPOPROTEIN YEAST.

FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 193 AA; 20921 MW; 46CFF28A948A00E CRC64;
Query Match 1.4%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 ALMSGC 23
Db 17 ALMSGC 23
RESULT 20
RISE ARCFU
ID RISE ARCFU STANDARD; PRT; 194 AA.
AC 027965;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L15e.
GN RPL15E OR AF2319.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klein H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyrpides N.C.,
RA Kirschmann R.D., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Kirsch E.F., Dougan J., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RT Nature 390:364-370(1997).
RN [2]
RP SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000944; AAB88937.1; -
CC PIR; G69539; G69539.
CC TIGR; AF2319; -
CC HAMAP; MF_00256; -; 1.
CC InterPro; IPR000439; Ribosomal_L15e.
CC Pfam; PF00827; Ribosomal_L15e; 1.
CC PROSITE; PS01194; RIBOSOMAL_L15E; 1.
CC Ribosomal protein; Complete proteome.
SQ SEQUENCE 194 AA; 23167 MW; 4B607B60F132BD23 CRC64;
Query Match 1.4%; Score 7; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 76 VDRDHPA 82
Db 134 VDRDHPA 140
RESULT 21

RPST MYCTU STANDARD; PRT; 212 AA.
 ID RPST MYCTU STANDARD; PRT; 212 AA.
 AC 050712;
 DT 01-NOV-1997 (Rel. 35, Last Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable RNA polymerase sigma-D factor.
 GN SIGD OR RV3414C OR MT3523 OR MTCY78.15.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolova M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weisman J., Khouri H., Gill J., Mikula A.,
 RA Bhalal W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z77165; CAB01009.1; -
 CC EMBL; AE007157; AAK47861.1; ALT_INIT.
 CC PIR; C70737; C70737.
 CC TIGR; MT3523; -
 CC Tuberculat; RV3414C; -
 CC Interpro; IPR000838; Sigma70_ECF.
 CC Pfam; PF04542; sigma70_r2; 1.
 CC Pfam; PF04545; sigma70_r4; 1.
 CC PROSITE; PS01063; SIGMA70_ECF; 1.
 CC Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
 CC DNA-binding; Complete proteome.
 CC
 CC POLYMERASE CORE BINDING (POTENTIAL).
 FT DNA BIND 75 88
 FT DNA BIND 176 195 H-T-H MOTIF (BY SIMILARITY).
 SQ SEQUENCE 212 AA; 22919 MW; 12A157F6F66B0C3 CRC64;

RESULT 22
 ID YIDX_ECOLI STANDARD; PRT; 218 AA.
 AC P31461; P76736;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yidx.
 GN YIDX OR B3696 OR SF3768.
 OS Escherichia coli, and
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NCBI_TaxID=562, 623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;
 RX MEDLINE=93315143; PubMed=7686882;
 RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
 RT "A genome sequence and analysis of 136 kilobases of the Escherichia coli
 RL genome: organizational symmetry around the origin of replication."
 RL Genomics 16:151-161(1993).
 RN [2]
 RP REVISIONS.
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157."
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [4]
 RP CONCEPTUAL TRANSLATION.
 RC SPECIES=E.coli;
 RA Rudd K.E.;
 RL Unpublished observations (JAN-2000).
 CC -1- CAUTION: Ref.1 and Ref.2 sequences differ from that shown due to a
 CC frameshift in position 188.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L10328; AAA62047.1; ALT_FRAME.
 CC EMBL; AE000446; AAC76719.1; ALT_FRAME.
 CC EMBL; AE015384; AAN45211.1; -
 CC Ecogene; Egi1719; yidx.
 CC Hypothetical protein; Transmembrane; Complete proteome.
 CC TRANSMEM 11 31 POTENTIAL.
 FT TRANSMEM 11 31
 SQ SEQUENCE 218 AA; 24169 MW; 73C7000506883D4 CRC64;

Query Match 488 LSADVA 494
 |||||
 DB 71 LSADVA 77
 Best Local Similarity 1.4%; Score 7; DB 1; Length 212;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 488 LSADVA 494
 |||||
 DB 71 LSADVA 77
 Best Local Similarity 1.4%; Score 7; DB 1; Length 218;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23
 DB 18 ALMLSGC 24

RESULT 23

TPIS_METER ID TPIS_METER STANDARD; PRT; 226 AA.
 AC 074025;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Triosephosphate isomerase (EC 5.3.1.1) (TIM).
 GN TPPI OR TPI.
 OS Methanobacterium bryantii.
 OC Archaea: Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanobacterium.
 OK NCBI_TaxID=2161;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 862;
 RA Schramm A.;
 RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RP CHARACTERIZATION.
 RX MEDLINE=96198612; PubMed=8925906;
 RA Kohlhoff M., Dahm A., Hensel R.;
 RT "Tetrameric triosephosphate isomerase from hyperthermophilic
 RT Archaea.";
 RL FEBS Lett. 383:245-250 (1996).
 CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycero-
 CC phosphate.
 CC -1- PATHWAY: Plays an important role in several metabolic pathways.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y11302; CAA72160.1; -.
 DR HAMAP: MF_00147; -; 1.
 DR InterPro: IPR003009; FMN_enzyme.
 DR InterPro: IPR000652; Triophos_isomase.
 DR Pfam: PF00121; TIM; 1.
 DR ProDom: PD001005; Triophos_isomase; 1.
 DR Tricore: TRIGR00419; tim; 1.
 DR PROSITE: PS00171; TIM; FALSE_NEG.
 KW Isomerases; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
 KW Pentose shunt.
 FT ACT SITE 97 97 BY SIMILARITY.
 SQ SEQUENCE 226 AA; 23907 MW; 3A6B9DB84A54E546 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 226;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 TGENALK 262
 DB 21 TGENALK 27

RESULT 24
 YX09_CAEEL STANDARD; PRT; 230 AA.
 AC Q1115;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypochemical 26.5 kDa protein C03B1.9 in chromosome X.
 GN C03B1.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OK NCBI_TaxID=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Martin J.;
 RN Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U40952; AAA81738.1; -.
 DR PIR: T15381; T15381.
 DR WormPep: C03B1.9; C03910.
 KW Hypochemical protein.
 SQ SEQUENCE 230 AA; 26473 MW; 082B700AB1229688 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 230;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 ITVNGNQ 455
 DB 58 ITVNGNQ 64

RESULT 25
 EF3_HUMAN STANDARD; PRT; 238 AA.
 AC P52757;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ephrin-A3 precursor (EPH-related receptor tyrosine kinase ligand 3)
 DE (LEK3-3) (EHK1 ligand) (EHK1-L).
 GN EFNA3 OR EPH3 OR LEK3 OR EFL-2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OK NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95140419; PubMed=7838529;
 RA Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,
 RA Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,
 RA Cerretti D.P., Beckmann M.P.;
 RT "Ligands for the receptor tyrosine kinases hek and elk: isolation of
 RT cDNAs encoding a family of proteins.";
 RL Oncogene 10:299-306 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95063919; PubMed=7973638;
 RA Davis S., Gale N.W., Aldrich T.H., Maisonnier P.C., Lhotak V.,
 RA Pawson T., Goldfarb M., Yancopoulos G.D.;
 RT "Ligands for EPH-related receptor tyrosine kinases that require
 RT membrane attachment or clustering for activity.";
 RL Science 266:816-819 (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ductum;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Mariani A., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loguettano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.M., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Gilwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, SKELETAL MUSCLE, SPLEEN,
 CC THYMUS, PROSTATE, TESTIS, OVARY, SMALL INTESTINE, AND PERIPHERAL
 CC BLOOD LEUKOCYTES.
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
 CC * This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U14187; AAC50078.1; -;
 DR EMBL: U13760; AAC52368.1; -;
 DR EMBL: BC017722; AAH17722.1; -;
 DR PIR: I38849; I38849.
 DR Genew: HGNC:3223; EFNA3.
 DR MIM: 601381; -;
 DR GO: GO:0005887; C:Integral to plasma membrane; TAS.
 DR GO: GO:0005005; F:transmembrane-ephrin receptor activity; TAS.
 DR GO: GO:0007287; P:cell-cell signaling; TAS.
 DR InterPro: IPR001799; Ephrin.
 DR Pfam: PF00812; Ephrin; 1.
 DR PRINTS: PRO1347; EPHRIN.
 DR ProDom: PD002533; Ephrin; 1.
 DR PROSITE: PS01299; EPHRIN; 1.
 KM Glycoprotein; GPI-anchor; Signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 238 EPHRIN-A3.
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 71 74 MISSING (IN REF. 2).
 SQ SEQUENCE 238 AA; 26350 MW; 8EPD6AE8FE33FDDA CRC64;
 Query Match 1.4%; Score 7; DB 1; Length 238;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 148 GPGGGA 154
 DB 78 GPGGGA 84
 RESULT 26
 YGIP YEAST STANDARD; PRT; 239 AA.
 AC P53223;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 27.6 kDa protein in RPL26B-ACB1 intergenic region.
 GN YG0303C.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=97435481; PubMed=9290212;
 RA Rieger M., Bruckner M., Schaefer M., Mueller-Auer S.,
 RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
 RT chromosome VII".
 RL Yeast 13:1077-1090(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC * This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z72821; CA97024.1; -;
 DR PIR: S64327; S64327.
 DR SGD: S0003268; CAX4.
 DR GO: GO:0030176; C:endoplasmic reticulum membrane, intrinsic P. .; IDA.
 DR GO: GO:0016462; F:pyrophosphatase activity; IDA.
 DR GO: GO:0006610; P:lipid biosynthesis; IGI.
 DR GO: GO:0006487; P:n-linked glycosylation; IMP.
 DR InterPro: IPR000326; PA_PTPase.
 DR Pfam: PF01569; PAP2; 1.
 DR SMART: SM00014; acidppc; 1.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 165 185 POTENTIAL.
 SQ SEQUENCE 239 AA; 27649 MW; C8EC49CBFC08F8F8 CRC64;
 Query Match 1.4%; Score 7; DB 1; Length 239;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 251 SVGALTG 257
 DB 169 SVGALTG 175
 RESULT 27
 FMDC METWP STANDARD; PRT; 272 AA.
 ID FMDC METWP
 AC Q31112;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Tungsten-containing formylmethanofuran dehydrogenase II subunit C
 DE (EC 1.2.99.5).
 GN FMDC.
 OS Methanococcus maripaludis.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanococcaceae; Methanococcus.
 OX NCBI_TaxID=39152;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JF;
 RA Yu J.-P., Whitman W.B.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO(2) AND
 CC METHANOFURAN (MFR) TO N-FORMYLMETHANOFURAN (CHO-MFR). THIS ENZYME
 CC IS OXYGEN-LABILE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Formylmethanofuran + H(2)O + acceptor = CO(2)
 CC + methanofuran + reduced acceptor.
 CC -1- COFACTOR: TUNGSTEN (BY SIMILARITY).
 CC -1- PATHWAY: Methanogenesis; first step.

```

CC -1- SUBUNIT: THIS ENZYME IS COMPOSED OF SEVEN SUBUNITS FWDA (65 kDa),
CC FWDB (53 kDa), FWDC (31 kDa), FWDD (15 kDa), FWDE, FWDF, AND FWDG
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FWDC/FWDC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF029842; AAB82591.1; -.
CC DR InterPro: IPR002489; DUF14.
CC DR Pfam: PF01493; GKGXG; 1.
CC DR Oxidoreductase: Tungsten; Methanogenesis; Repeat.
CC FT DOMAIN 77 210 7 X 13 AA REPEATS OF [GW]-X-X-[MLP]-X-X-
CC G-X-[IL]-X-X-[IV]-X-X-G.
CC FT REPEAT 77 89 1.
CC FT REPEAT 96 108 2.
CC FT REPEAT 115 127 3.
CC FT REPEAT 141 153 4.
CC FT REPEAT 160 172 5.
CC FT REPEAT 179 191 6.
CC FT REPEAT 198 210 7.
CC SQ SEQUENCE 272 AA; 29174 MW; 50912BD8B47A4BF0 CRC64;

Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 272;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKITVNG 453
DB 102 GKITVNG 108

RESULT 28
SYGA COXBU STANDARD; PRT; 319 AA.
ID SYGA COXBU
AC P946T6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glycyl-tRNA synthetase alpha chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE alpha chain) (GLYRS).
GN GLYQ OR CBU913.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_Taxid=777;
OC [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RA Williams H., Jaeger C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RA Willems H., Jaeger C.;
RT Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RT [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RA MEDLINE=22608657; PubMed=12704232;
RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tettelin H., Davidson T.M., Beanan M.J.,
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouli H.M., Lee K.H., Carty H.A., Scanlan D., Hainzen R.A.,
RA Thompson H.A., Samuel J.E., Fraser C.W., Heidelberg J.F.;
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
RT burnetii."
RT Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyl-tRNA(Gly).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

```

```

CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y10435; CAA71456.1; -.
CC DR EMBL: AE016966; AA091404.1; -.
CC DR TIGR; CB01913; -.
CC DR HAMAP; MF_00254; -.
CC DR InterPro: IPR002310; tRNA synth 2e.
CC DR InterPro: IPR006194; tRNA synth Gly.
CC DR Pfam: PF02091; tRNA synth 2e; 1.
CC DR PRINTS; PR01044; TRNASYNTHGA.
CC DR PRODOM; PD006985; tRNA synth 2e; 1.
CC DR TIGRFAMs; TIGR00388; glyo; 1.
CC DR PROSITE; PS50861; AA tRNA LIGASE II GYAB; 1.
CC KM Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding.
CC FT DOMAIN 301 305 305
CC SQ SEQUENCE 319 AA; 36397 MW; FF52579A9F0787F8 CRC64;

Query Match
Best Local Similarity 1.4%; Score 7; DB 1; Length 319;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 LQPFDM 236
DB 28 LQPFDM 34

RESULT 29
HAM1 STRP3 STANDARD; PRT; 328 AA.
ID HAM1 STRP3
AC O8K8I7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE HAM1 protein homolog.
GN SPYM3_0263 OR SP51596.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=198466;
OC [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,
RA Yamagaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,
RA Hayashi H., Hamada S.;
RT "The genome of invasive Streptococcus pyogenes; a comparative analysis
RT of S. pyogenes SSI-1, SF370 and MGAS832."
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE HAM1 STRPAB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE014141; AAM7870.1; -
DR EMBL; AP005146; BAC64691.1; -
DR HAMAP; MF 01405; fused; 1.
DR InterPro; IPR002637; Hamip_1like.
DR Pfam; PF01725; Hamip_1like; 1.
DR ProDom; PD004952; Hamip_1like; 1.
DR TIGRFAMs; TIGR00042; TIGR00042; 1.
KM Hydrolyase; Complete proteome.
FT DOMAIN 1 129 UNKNOWN.
SQ SEQUENCE 328 AA; 36395 MW; 91A6D54B6A08888 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 AVYLPFG 372
DB 105 AVYLPFG 111

RESULT 30
HAM1_STRP8 STANDARD; PRT; 328 AA.
ID HAM1_STRP8
AC Q8P2D2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE HAM1 protein homolog.
GN SPY0362.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RA MedLine=21927593; PubMed=11917108;
RA Smoot J.C., Barlian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Studevant D.E., Ricklefs S.W., Porcella S.P.,
RA Pakins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
CC -1- SIMILARITY: BELONGS TO THE HAM1 NTPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 AVYLPFG 372
DB 105 AVYLPFG 111

RESULT 31
HAM1_STRPY STANDARD; PRT; 328 AA.
ID HAM1_STRPY
AC G9A1B6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE HAM1 protein homolog.
GN SPY0362.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RA MedLine=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian X., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yvan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -1- SIMILARITY: BELONGS TO THE HAM1 NTPASE FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE006500; AAC3407.1; -
DR HAMAP; MF 01405; fused; 1.
DR InterPro; IPR002637; Hamip_1like.
DR Pfam; PF01725; Hamip_1like; 1.
DR ProDom; PD004952; Hamip_1like; 1.
DR TIGRFAMs; TIGR00042; TIGR00042; 1.
KM Hydrolyase; Complete proteome.
FT DOMAIN 1 129 UNKNOWN.
SQ SEQUENCE 328 AA; 36292 MW; BA7A3677BCAFA254F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 AVYLPFG 372
DB 105 AVYLPFG 111

RESULT 32
PYRB_LACLE STANDARD; PRT; 351 AA.
ID PYRB_LACLE
AC Q60257;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartate carbamoyltransferase (EC 2.1.3.2) (Aspartate
DE transcarbamylase) (ATCase).
GN PYRB.
OS Lactobacillus leichmannii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;


```

OC Lactobacillus.
OX NCBI_TaxId=28039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 20076;
RA MEDLINE=96285745; PubMed=8725005;
RT Becker J., Brendel M.;
RT "Molecular cloning and characterization of the pyrB gene of
RT Lactobacillus leichmannii encoding aspartate transcarbamylase.";
RL Biochimie 78:3-3(1996).
CC -I- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
CC + N-carbamoyl-L-aspartate.
CC -I- PATHWAY: Pyrimidine biosynthesis; second step.
CC -I- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X84282; CAA59021.1; -
CC EMBL: X78999; CAA55634.1; -
CC PIR: T46956; T46956.
CC HSSP: P00479; 3CSU.
CC HAMAP: MF_00001; -1.
CC InterPro: IPR006130; Asp/Om_Cotranf.
CC InterPro: IPR002082; Asp_CarbmLtransf.
CC InterPro: IPR006131; OTCase_O.
CC InterPro: IPR006132; OTCase_P.
CC Pfam: PF00185; OTCase; 1.
CC Pfam: PF02729; OTCase; N: 1.
CC PRINTS: PR00100; AOTCASE.
CC TIGRPFAM: TIGR00670; asp_carb_tr; 1.
CC PROSITE: PS00097; CARBAMOYLTRANSFERRASE; 1.
CC Pyrimidine biosynthesis; Transferrase.
CC KM SEQUENCE 351 AA; 39857 MW; 8C36C4B1AEBD956 CRC64;
SQ
Query Match 1.4%; Score 7; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 487 QLSADVV 493
46 QLSADVV 52
RESULT 33
AVRC_PSESG STANDARD; PRT; 352 AA.
ID AVRC_PSESG STANDARD; PRT; 352 AA.
AC P13836;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Avirulence C protein.
GN AVRC.
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxId=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Race 0;
RC MEDLINE=89008107; PubMed=3049552;
RA Tamaki S., Dahlbeck D., Staakawicz B., Keen N.T.;
RT "Characterization and expression of two avirulence genes cloned from
RT Pseudomonas syringae pv. glycinea.";
RL J. Bacteriol. 170:4846-4854(1988).
CC -I- DISEASE: THE TERM AVIRULENT DESCRIBES A POTENTIALLY VIRULENT
CC PATHOGEN WHICH IS UNABLE TO INDUCE DISEASE SYMPTOMS IN SPECIFIC
CC CULTIVARS OF SOYBEAN.

```

```

CC -I- SIMILARITY: 46% IDENTITY TO AVIRULENCE B PROTEIN OF THE SAME
CC STRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M22219; AAA88428.1; -
CC PIR: B43649; B43649.
CC KM Virulence.
CC SQ SEQUENCE 352 AA; 39147 MW; B922BF6F41A77A97 CRC64;
SQ
Query Match 1.4%; Score 7; DB 1; Length 352;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 364 TDAVYLP 370
249 TDAVYLP 255
RESULT 34
METL_PEA STANDARD; PRT; 374 AA.
ID METL_PEA STANDARD; PRT; 374 AA.
AC P49613;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-adenosylmethionine synthetase 2 (EC 2.5.1.6) (Methionine
DE adenosyltransferase 2) (Adomet synthetase 2).
GN SAM52.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euphorbia; Fabales; Fabaceae; Papilionoideae; Viciales; Pisum.
OX NCBI_TaxId=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska;
RA Gomez L., Carrasco P.;
RL Submitted (OCT-1984) to the EMBL/Genbank/DDBJ databases.
CC -I- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC METHIONINE AND ATP.
CC -I- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC diphosphate + S-adenosyl-L-methionine.
CC -I- COFACTOR: Binds 2 divalent ions, such as magnesium or cobalt, and
CC 1 potassium ion per subunit (by similarity).
CC -I- PATHWAY: Activated methyl cycle.
CC -I- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X82077; CAA57581.1; -
CC EMBL: U36681; AAA88773.1; -
CC PIR: S66352; S66352.
CC HSSP: P04384; 1MXB.
CC InterPro: IPR002133; S-Adomet_synt.
CC Pfam: PF00438; S-Adomet_synt; 1.
CC Pfam: PF02772; S-Adomet_synt2; 1.
CC Pfam: PF02773; S-Adomet_synt3; 1.
CC TIGRPFAM: TIGR01034; metK; 1.
CC PROSITE: PS00376; ADOMET_SYNTHETASE_1; 1.
CC PROSITE: PS00377; ADOMET_SYNTHETASE_2; 1.

```

KM Transferrase; One-carbon metabolism; Multigene family; ATP-binding;
 KM Magnesium; Potassium; Metal-binding.
 FT NP BIND 121 126 ATP (POTENTIAL).
 FT METAL 19 19 MAGNESIUM (BY SIMILARITY).
 FT METAL 45 45 POTASSIUM (BY SIMILARITY).
 FT METAL 273 273 POTASSIUM (BY SIMILARITY).
 FT METAL 281 281 MAGNESIUM (BY SIMILARITY).
 FT BINDING 149 149 ATP (POTENTIAL).
 SO SEQUENCE 374 AA; 40976 MW; 90921A5AE7420882 CRC64;
 Query Match 1.4%; Score 7; DB 1; Length 374;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 439 SIVANGL 445
 DB 294 SIVANGL 300
 RESULT 35
 ID_NIR_RHOSH STANDARD; PRT; 374 AA.
 AC 053239;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
 GN NIKR.
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Rhodospirillaceae; Rhodospirillum.
 OX NCBI_TaxID=1063;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=2.4.3;
 RX MEDLINE=91715533; PubMed=9023188;
 RA Toques I.E., Kwiatkowski A.V., Shi J., Shapleigh J.P.;
 RT "Characterization and regulation of the gene encoding nitrite
 reductase in Rhodospirillum rubrum 2.4.3.";
 RJ J. Bacteriol. 179:1090-1095(1997).
 CC -1- CATALYTIC ACTIVITY: Nitrite oxidoreductase (NIR) =
 nitrite + ferrioxochrome c.
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE
 1 COPPER. FAD. TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER.
 CC WHILE TYPE 2 COPPER IS HELD BY RESIDUES FROM EACH OF TWO MONOMERS
 OF THE TRIMER. NITRITE IS BOUND TO THE TYPE 2 COPPER SITE.
 CC PSEUDODAZURIN IS THE PHYSIOLOGICAL ELECTRON DONOR FOR THE CU-NIR IN
 VITRO (BY SIMILARITY).
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
 CC -1- DOMAIN: THE TYPE 1 COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
 ELECTRON TRANSFER FROM PSEUDODAZURIN TO THE TYPE 1 COPPER SITE
 OF NIR, WHICH COMPREHENDS THE CATALYTIC CENTER OF NIR FOR THE
 REDUCTION OF NITRITE.
 CC -1- SIMILARITY: Contains 2 plastocyanin-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; U62291; AAB05767.1; -.
 DR HSSP; P25006; INIF.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR001287; CNO2_reductase.
 DR Pfam; PF00394; Cu-oxidase; 2.
 DR PRINTS; PR00695; CUNO2RDRTASE.
 KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
 Nitrates assimilation; Repeat; Periplasmic; Signal.

FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 374 COPPER-CONTAINING NITRITE REDUCTASE.
 FT DOMAIN 93 189 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 254 355 PLASTOCYANIN-LIKE 2.
 FT METAL 126 126 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 131 131 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 166 166 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 167 167 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 177 177 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 182 182 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 338 338 COPPER (TYPE 2) (BY SIMILARITY).
 SO SEQUENCE 374 AA; 40308 MW; 3406B5B5E7DD9934 CRC64;
 Query Match 1.4%; Score 7; DB 1; Length 374;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 142 DFHAATG 148
 DB 129 DFHAATG 135
 RESULT 36
 ID_NIR_ALCPA STANDARD; PRT; 376 AA.
 AC P38501;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Copper-containing nitrite reductase precursor (EC 1.7.2.1) (Cu-NIR).
 GN NIKR OR NIR.
 OS Alcaligenes faecalis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Alcaligenes.
 OX NCBI_TaxID=511;
 RN [1]
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 34-48 AND 192-240.
 RC STRAIN=S-6;
 RX MEDLINE=93294530; PubMed=8515232;
 RA Nishiyama M., Suzuki J., Kukimoto M., Ohnuki T., Horinouchi S.,
 RA Beppu T.;
 RT "Cloning and characterization of a nitrite reductase gene from
 Alcaligenes faecalis and its expression in Escherichia coli.";
 RJ J. Gen. Microbiol. 139:725-733(1993).
 RN [2]
 RC X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND MUTAGENESIS.
 RC STRAIN=S-6;
 RX MEDLINE=94227056; PubMed=8172899;
 RA Kukimoto M., Nishiyama M., Murphy M.E.P., Turley S., Adman E.T.,
 RA Horinouchi S., Beppu T.;
 RT "X-ray structure and site-directed mutagenesis of a nitrite reductase
 from Alcaligenes faecalis S-6: roles of two copper atoms in nitrite
 reduction.";
 RJ Biochemistry 33:5246-5252(1994).
 RN [3]
 RC X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=96032602; PubMed=7547950;
 RA Murphy M.E., Turley S., Kukimoto M., Nishiyama M., Horinouchi S.,
 RA Sasaki H., Tanokura M., Adman E.T.;
 RT "Structure of Alcaligenes faecalis nitrite reductase and a copper
 site mutant, M50E, that contains zinc.";
 RJ Biochemistry 34:12107-12117(1995).
 RN [4]
 RC X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=98019216; PubMed=9353305;
 RA Murphy M.E., Turley S., Adman E.T.;
 RT "Structure of nitrite bound to copper-containing nitrite reductase
 from Alcaligenes faecalis. Mechanistic implications.";
 RJ J. Biol. Chem. 273:28455-28460(1997).
 CC -1- CATALYTIC ACTIVITY: Nitrite oxidoreductase (NIR) =
 nitrite + ferrioxochrome c.
 CC -1- COFACTOR: CONTAINS ONE ATOM OF TYPE 1 COPPER AND ONE ATOM OF TYPE
 1 COPPER. FAD. TYPE 1 COPPER IS BOUND WITHIN A SINGLE MONOMER.

RA Chen J.-Y., Chang W.-C., Chang T., Chang W.-C., Liu M.-Y., Payne W.J.,
RA le Gall J.;
RT "Cloning, characterization, and expression of the nitric oxide-
RT generating nitrite reductase and of the blue copper protein genes of
RT *Achromobacter cycloclastes*.";
RL Biochem. Biophys. Res. Commun. 219:423-428(1996).
RN [2]
RP SEQUENCE OF 39-378.
RC STRAIN=IAM 1013;
RX MEDLINE=91308101; PubMed=1830217;
RA Fenderson F.F., Kumar S., Adman E.T., Liu M.-Y., Payne W.J.,
RA le Gall J.;
RT "Amino acid sequence of nitrite reductase: a copper protein from
RT *Achromobacter cycloclastes*.";
RL Biochemistry 30:7180-7185(1991).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=91320115; PubMed=1862344;
RA Godden J.W., Turley S., Teller D.C., Adman E.T., Liu M.-Y.,
RA Payne W.J., le Gall J.;
RT "The 2.3-A X-ray structure of nitrite reductase from *Achromobacter*
RT *cycloclastes*.";
RL Science 253:438-442(1991).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96070866; PubMed=749203;
RA Adman E.T., Godden J.W., Turley S.;
RT "The structure of copper-nitrite reductase from *Achromobacter*
RT *cycloclastes* at five pH values, with NO₂-bound and with type II
RT copper depleted.";
RL J. Biol. Chem. 270:27458-27474(1995).
CC -I- CATALYTIC ACTIVITY: Nitric oxide + H(2)O + ferriocytochrome c =
CC nitrite + ferriocytochrome c.
CC -I- COPFACTOR: CONTAINS ONE ATOM OF TYPE I COPPER AND ONE ATOM OF TYPE
CC II COPPER; FAD.
CC -I- PATHWAY: Nitrate assimilation (denitrification).
CC -I- SUBUNIT: Homotrimer.
CC -I- SUBCELLULAR LOCATION: Periplasmic.
CC -I- DOMAIN: THE TYPE I COPPER SITE IN NIR PLAYS A CRUCIAL ROLE FOR
CC ELECTRON TRANSFER FROM PSEUDAZURIN TO THE TYPE II COPPER SITE
CC OF NIR, WHICH COMPRISES THE CATALYTIC CENTER OF NIR FOR THE
CC REDUCTION OF NITRITE.
CC -I- SIMILARITY: Contains 2 plastocyanin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48635; CA88564.1; -.
DR PIR; JC4648; JC4648.
DR PDB; 2NRD; 07-DEC-95.
DR PDB; 1NIA; 07-DEC-95.
DR PDB; 1NIB; 07-DEC-95.
DR PDB; 1NIC; 07-DEC-95.
DR PDB; 1NID; 07-DEC-95.
DR PDB; 1NIE; 07-DEC-95.
DR PDB; 1NIF; 07-DEC-95.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR001287; CuNO₂-reductase.
DR InterPro; IPR006311; Tat.
DR Pfam; PF00394; Cu-oxidase; 2.
DR PRINTS; PR00695; CUINO2RD1ASE.
DR TIGRFAMs; TIGR01409; Tat signal seq; 1.
KW Oxidoreductase; Copper; Metal-binding; Flavoprotein; FAD;
KW Nitrate assimilation; Repeat; Periplasmic; Signal; 3D-structure.
FT SIGNAL 1 38
FT CHAIN 39 378 COPPER-CONTAINING NITRITE REDUCTASE.
FT DOMAIN 39 213 PLASTOCYANIN-LIKE 1.
FT DOMAIN 214 378 PLASTOCYANIN-LIKE 2.

FT METAL 133 133 COPPER (TYPE 1).
FT METAL 138 138 COPPER (TYPE 2).
FT METAL 173 173 COPPER (TYPE 2).
FT METAL 174 174 COPPER (TYPE 1).
FT METAL 183 183 COPPER (TYPE 1).
FT METAL 188 188 COPPER (TYPE 1).
FT METAL 344 344 COPPER (TYPE 2).
FT METAL 344 344
FT METAL 47 49
FT STRAND 52 54
FT TURN 60 61
FT STRAND 69 69
FT STRAND 76 76
FT TURN 89 89
FT TURN 92 93
FT STRAND 96 102
FT TURN 103 104
FT STRAND 105 105
FT STRAND 110 110
FT TURN 115 116
FT STRAND 118 125
FT TURN 127 128
FT STRAND 133 133
FT STRAND 136 137
FT TURN 138 139
FT HELIX 143 149
FT STRAND 152 152
FT TURN 154 155
FT STRAND 156 163
FT STRAND 168 173
FT TURN 177 178
FT HELIX 180 184
FT TURN 185 187
FT STRAND 189 195
FT TURN 197 197
FT STRAND 198 198
FT STRAND 200 201
FT TURN 203 204
FT STRAND 207 208
FT STRAND 212 221
FT STRAND 225 225
FT TURN 227 228
FT STRAND 231 231
FT HELIX 237 248
FT TURN 249 250
FT STRAND 255 258
FT TURN 259 260
FT STRAND 261 261
FT TURN 262 265
FT HELIX 267 269
FT STRAND 271 274
FT TURN 275 276
FT STRAND 278 285
FT STRAND 290 290
FT STRAND 292 295
FT TURN 296 296
FT STRAND 299 299
FT TURN 304 305
FT STRAND 308 309
FT TURN 313 316
FT STRAND 321 321
FT STRAND 323 324
FT STRAND 325 332
FT STRAND 337 343
FT STRAND 346 350
FT TURN 351 352
FT STRAND 355 361
FT TURN 366 368
FT TURN 368 368
SQ SEQUENCE 378 AA; 40771 MW; A70B52B814090EA5 CRC64;
Query Match 1.4%; Score 7; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;
QY 142 DFHAATG 148

Db 136 DPHATG 142

RESULT 38
ID YM12_MYCTU STANDARD; PRT; 378 AA.AC Q10400;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein Rv2212
GN Rv2212 OR MT2268 OR MTCY190.23.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Ospreme J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterlind T., Weidman J., Khouri H., Gail J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.-1- SIMILARITY: Belongs to the adenyl cyclase class-4/guanylyl
cyclase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).CC
CC EMBL; Z70283; CAA94255.1; -
DR EMBL; AE007072; AAK46554.1; ALT_INIT.
DR PIR; E70786; E70786.
DR HSP; P19754; LAMK.
DR TIGR; MT2268; -
DR Tuberculin; Rv2212; -
DR InterPro; IPR001054; G_cyclase.
DR Pfam; PF00211; guanylate_cyc; 1.
DR SMART; SM00044; CYCG; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 208 317 GUANYLATE CYCLASE.
SQ SEQUENCE 378 AA; 39775 MW; AF26A9CFC13C91DC CRC64;Query Match 1.4%; Score 7; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 488 LSADVA 494

Db 74 LSADVA 80

RESULT 39
ID REC2_LACLA STANDARD; PRT; 387 AA.AC Q01840; O9C1K3;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RecA protein, chromosomal (Recombinase A).
GN RECA OR IL0354.
OS Lactococcus lactis (subsp. *lactis*) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ML3;
RX MEDLINE=92384590; PubMed=1514816;
RA Duwat P., Ehrlich S.D., Grues A.;
RT "Use of degenerate primers for polymerase chain reaction cloning and
RT sequencing of the Lactococcus lactis subsp. *lactis* recA gene.";
RL Appl. Environ. Microbiol. 58:2674-2678 (1992).[2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT *lactis* ssp. *lactis* IL1403.";
RL Genome Res. 11:731-753 (2001).-1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
SINGLE-STRANDED DNA. THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
DNA BY DUPLICATING DNA. AND THE ATP-DEPENDENT HYBRIDIZATION OF
HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the recA family.CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).CC EMBL; M88106; AAK25216.1; -
DR EMBL; AE006272; AAK04452.1; -
DR PIR; B86669; B86669.
DR HSP; P26345; IG19.
DR HAMAP; MF_00268; -; 1.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001553; RecA.
DR Pfam; PR00154; recA; 1.
DR PRINTS; PR00142; RECA.
DR ProDom; PD000229; RECA; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00321; RECA_1; 1.
DR PROSITE; PS50162; RECA_2; 1.
DR PROSITE; PS50163; RECA_3; 1.
KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding;
KW Complete proteome.FT NP BIND 80 87 ATP (BY SIMILARITY).
FT NP BIND 80 87 R -> S (IN REF. 1).
FT CONFLICT 298 298 D -> E (IN REF. 1).
FT CONFLICT 298 298 E -> DV (IN REF. 1).
FT CONFLICT 332 333 T -> A (IN REF. 1).
FT CONFLICT 343 343 EA -> DS (IN REF. 1).
FT CONFLICT 350 351 TEDSTKATKATKKEKEVEETIEIELELED -> EEEETTA
FT CONFLICT 357 387 FGN (IN REF. 1).

SQ SEQUENCE 387 AA; 41477 MM; A8CAFC0BB27BF14F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 387;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 YGEGQLQ 231

DB 135 YGEGQLQ 141

QY 439 SIVANGL 445

DB 292 SIVANGL 298

Search completed: August 27, 2003, 18:43:03
Job time : 46 secs

RESULT 40

METK CATRO

ID METK CATRO STANDARD; PRT; 393 AA.

AC 096551;

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE S-adenosylmethionine synthetase 1 (Sc 2.5.1.6) (methionine

adenosyltransferase 1) (Adomet synthetase 1).

GN SAMS1.

OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; Lamiales; Gentianales; Apocynaceae; Rauvolfiaceae; Vincet;

OC Catharanthus.

OX NCBI_Taxid=4058;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97188555; PubMed=9037140;

RA Schroeder G., Eichel J., Breinig S., Schroeder J.;

RT "Three differentially expressed S-adenosylmethionine synthetases from

Catharanthus roseus: molecular and functional characterization.";

Plant Mol. Biol. 33:211-222(1997).

CC -!- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM

METHIONINE AND ATP.

CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +

diphosphate + S-adenosyl-L-methionine.

CC -!- COFACTOR: Binds 2 divalent ions, such as magnesium or cobalt, and

1 potassium ion per subunit (By similarity).

CC -!- PATHWAY: Activated methyl cycle.

CC -!- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: 271271; CA95856.1; -.

DR HSSP: P04384; IMXB

DR InterPro: IPR002133; S-Adomet_synt.

DR Pfam: PF00438; S-Adomet_synt; 1.

DR Pfam: PF02772; S-Adomet_syntD2; 1.

DR Pfam: PF02773; S-Adomet_syntD3; 1.

DR TIGRFAMs: TIGR01034; metK; 1.

DR PROSITE: PS00376; ADOMET SYNTHETASE_1; 1.

DR PROSITE: PS00377; ADOMET SYNTHETASE_2; 1.

KW Transferrase; One-carbon metabolism; Multigene family; ATP-binding;

KW Magnesium; Potassium; Metal-binding.

FT NP BIND 119 124 ATP (POTENTIAL).

FT METAL 17 17 MAGNESIUM (BY SIMILARITY).

FT METAL 43 43 POTASSIUM (BY SIMILARITY).

FT METAL 271 271 POTASSIUM (BY SIMILARITY).

FT METAL 279 279 MAGNESIUM (BY SIMILARITY).

FT BINDING 147 147 ATP (POTENTIAL).

SQ SEQUENCE 393 AA; 43050 MM; 04371F2B55BE386F CRC64;

Query Match 1.4%; Score 7; DB 1; Length 393;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:26:57 ; Search time 102 Seconds

(without alignments)
1270.024 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626
Sequence: 1 MSKPTLITKTLICALSALML.....NKGGQLSADPVAKAKKXPN 502

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SPREMBL_23.*

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_oranella:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1487.5	56.6	510	16	Q8XPY3 ralstonia s
2	1388	52.9	390	16	Q9JYE1 neisseria m
3	1386	52.8	386	16	Q9JTB8 neisseria m
4	787	30.0	486	2	Q8KX44 hyphomicrob
5	751.5	28.6	361	1	Q9P9H9 haloarcula
6	481	18.3	360	2	Q9AOC7
7	451	17.2	360	2	Q68601
8	442	16.8	455	16	Q8U7R7
9	439.5	16.7	376	16	Q922Z9
10	431	16.4	376	16	Q8YB82
11	429.5	16.4	364	2	Q31380
12	429	16.3	376	16	Q8FX29
13	390.5	14.9	478	16	Q8FTV6
14	258	9.8	343	16	Q8FSM1
15	253	9.6	179	2	Q9JTN1
16	250.5	9.5	379	17	Q9HOF4

17	248.5	9.5	338	16	Q8YQ92	Q8YQ92 anabaena sp
18	236	9.0	179	2	Q8RLT4	Q8RLT4 mesorhizobi
19	233	8.9	179	2	Q8RLT4	Q8RLT4 ensifer sp.
20	232	8.8	179	2	Q8RLT4	Q8RLT4 ensifer sp.
21	230.5	8.8	180	2	Q9JTN6	Q9JTN6 nitrosomona
22	228.5	8.7	180	2	Q9JTN3	Q9JTN3 nitrosomona
23	228.5	8.7	180	2	Q9JTN4	Q9JTN4 nitrosomona
24	228	8.7	179	2	Q8RLS6	Q8RLS6 pseudomonas
25	226.5	8.6	180	2	Q9JTN2	Q9JTN2 nitrosomona
26	224.5	8.5	172	2	Q8KPM7	Q8KPM7 uncultured
27	223	8.5	179	2	Q8RLT1	Q8RLT1 ochrobactru
28	222	8.5	138	16	Q8K098	Q8K098 neisseria m
29	222	8.5	163	16	Q9JVA7	Q9JVA7 neisseria m
30	221	8.4	178	2	Q9JTN5	Q9JTN5 ochrobactru
31	219.5	8.4	180	2	Q9JTN5	Q9JTN5 nitrosomona
32	219	8.3	179	2	Q8RLS8	Q8RLS8 ochrobactru
33	219	8.3	180	2	Q8RLT5	Q8RLT5 azospirillum
34	216.5	8.2	172	2	Q8KPL0	Q8KPL0 uncultured
35	216	8.2	179	2	Q8RLT0	Q8RLT0 ochrobactru
36	215	8.2	179	2	Q8KPM6	Q8KPM6 uncultured
37	213.5	8.1	172	2	Q8KPM5	Q8KPM5 uncultured
38	211.5	8.1	172	2	Q8KPM5	Q8KPM5 uncultured
39	211.5	8.1	172	2	Q8KPM9	Q8KPM9 uncultured
40	211.5	8.1	172	2	Q8KPM9	Q8KPM9 uncultured
41	211.5	8.1	172	2	Q8KPM9	Q8KPM9 uncultured
42	210.5	8.0	172	2	Q8KPM8	Q8KPM8 uncultured
43	210.5	8.0	172	2	Q8KPM8	Q8KPM8 uncultured
44	210.5	8.0	172	2	Q8KPM8	Q8KPM8 uncultured
45	209.5	8.0	172	2	Q8KPM5	Q8KPM5 uncultured

ALIGNMENTS

RESULT 1
Q8XPY3 PRELIMINARY; PRT; 510 AA.
AC Q8XPY3;
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 23, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Probable major ANAEROBICALLY induced outer membrane
DE protein (EC 1.7.99.3).
DE RSP1503 OR R503038.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Broctier P., Camus J.C., Catecolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Denape N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigler P., Thebault P., Whalen M., Wincker P., Levy M.,
RT Weisenbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
DR EMBL: AL646085; CAD18654.1;
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR001287; CuNO2_reductase.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR003088; CytC.
DR Pfam: PF00394; Cytochrome c; 1.
DR Pfam: PF00034; Cytochrome c; 1.
DR PRINTS: PR00695; CUNO2RPTASE.
DR PROSITE: PS00190; CYTOCHROME C; 1.
KW Oxidoreductase; Plasmid; Complete proteome.
SQ SEQUENCE 510 AA; 54600 MW; 8008105DD99459AC CRC64;

Query Match	56.6%	Score 1487.5	DB 16	Length 510
Best Local Similarity	62.4%	Pred. No. 3e-92		
Matches 277	Conservative	63	Mismatches 99	Indels 5
			Gaps	4
Qy	62	IDAIVTHAPEVPPPRDRDPDRPAKVYVKKMEVKEKVLADVEYQFTFGGVGQGMIRARE	121	
Db	59	IRAVITSPNNVPPFPHRNYPKAVIYELEIEKEMISSEGVSTFTPTFGTIVGASTIRARQ	118	
Qy	122	GDITVEQPSNHPDSKMPHNVDFFHAATGPGGAEASFTAPGHTSTSPFALQGLVYHCA	181	
Db	119	GDITVEFLKHPHDSKMPHNIDHGVTPGCGGAASFTAPGHSQFTFKALNGVYVYHCA	178	
Qy	182	VAPVGMHIANMGYGLIYEPKKGELPKVDKEYYVMQGFPTTKKNGEGLQIPDMDEKAIRE	241	
Db	179	TAPVGMHIANMGYGLIYEPKKGELPKVDKEYYVMQGFPTTKKNGEGLQIPDMDEKAIDE	238	
Qy	242	DAEYVFNFGSVALTGEMNAKKKVEETRLPFVNGGPNLTSSEFHYIGELFDDKVEHGGKG	301	
Db	239	RPTVYLFNGABGALGDKAMHAKTGETVRIFFVNGGPNLVSSFFHYIGALFDDVRIEGGTN	298	
Qy	302	ENHNIIQTLLIPAGGAITFEKVDYGVGYLVYDAIFRAFNKALGILKVEGSENHETYSH	361	
Db	299	VQKAVQTLLIPAGGAIVVFTFARVGSYLVVDHSHFRANKGAMAIIKIDGEDKPIYSG	358	
Qy	362	KQTDPAVYIPEGA-FOAIDTQEAPEKTPAPAMT--QEOIKAKGATYSNSCAACHQPGKGVP	418	
Db	359	KELDAVYIGDRAGPNLSATVRAQAASAGTSLSMODOVQGRALFAGTCSVCHQSGNTGIP	418	
Qy	419	NAFPLASDYLNAADHAPASIVVANGLSGKITVNGNGVSWPAPLA-LSDDOIANVITVT	477	
Db	419	GVFPLASDYLNAADHAPASIVVANGLSGKITVNGNGVSWPAPLA-LSDDOIANVITVT	478	
Qy	478	LSNFGNKGQSLSADDDAVAKAKTKRP	501	
Db	479	LNSMNPGRGVAEDV-KTIRARQ	501	

RESULT 2

ID	Q9JYE1	PRELIMINARY	PRT	390 AA.
AC	Q9JYE1			
DT	01-OCT-2000	(TREMBLrel. 15, Created)		
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)		
DT	01-OCT-2002	(TREMBLrel. 22, Last annotation update)		
DE	Major anaerobically induced outer membrane protein.			
GN	MMB1623.			
OS	Neisseria meningitidis (serogroup B).			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;			
OC	Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=491;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MC58 / Serogroup B;			
RC	MEDLINE=2017575; PubMed10710307;			
RA	Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,			
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,			
RA	Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,			
RA	Hate D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,			
RA	Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,			
RA	Cotton M.D., Utteback T.R., Khouri H., Qin H., Vamathevan J.,			
RA	Gill J., Scariato V., Masiagnan V., Piazza M., Grandi G., Sun L.,			
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,			
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain			
RL	MC58."			
RL	Science 287:1809-1815 (2000).			
DR	EMBL; AE002512; AAF41975.1; .			
DR	HSSP; P25006; INF.			
DR	TIGR; NMB1623; .			
DR	InterPro; IPR001117; Cu-oxidase.			
DR	InterPro; IPR001287; CUNO2_reductase.			
DR	Pfam; PF00394; Cu-oxidase; 2.			
DR	PRINTS; PR00695; CUNO2RDTASE.			
KW	Complete proteome.			

SEQ	SEQUENCE	390 AA;	40763 MW;	CS03F9D47DD169D CRC64;
QY	Query Match	52.9%;	Score 1388;	DB 16; Length 390;
Db	Best Local Similarity	70.0%;	Pred. No. 1,1e-85;	
Matches	266;	Conservative 40;	Mismatches 66;	Indels 8; Gaps 3;
QY	11 LICALSALMSGCSNQADKKAQPPSSSYVDAAKTA-NADNMAISOEGSELVIDAIVTHA	69		
Db	9 MIALSFALMACG----GEPAAQAPAEETFPAAAEASSAAQTPAETPSELFPVIDAVTHA	64		
QY	70 PEVPPVDRDPAKVVVMETVEKVMRLADSEVQFMFGQVGGOMIRVREGDTIEVOF	129		
Db	65 PEVPALDRDPAKVRVMETVEKTMMEDEVERVYMTFDDVDGCMIRVREGDTIEVEF	124		
QY	130 SNHDSKMPHNVDFFHATGPGGGAESFTADGHTSTFSFKALQGLVYTHCAVAPVGMH	189		
Db	125 SNPSSIVPHNVDPHATGQGGGAATFTAGRTSTFSFKALQGLYTHCAVAPVGMH	184		
QY	190 ANGMVGLILVEPKGLPVNDKVEYVVMQDFTTKGKYEGQGIQPFDMKAIEDDAEYVEN	249		
Db	185 ANGMVGLILVBPKEGLPVNDKEFYIVQDFTYKKGKGAQGIQPFDMDAVAEAPYVEN	244		
QY	250 GSVGLTGENMLKAKVGETVRLPYNGGPGNLTSSPHVAGLIFDKVHFEQGGENHIQTT	309		
Db	245 GHVGAIADNMLKAKAGETVMYVNGGPNLVSSPHVAGLIFDKVYVGGGLINENYST	304		
QY	310 LIPAGGAITEFPKVDVPGDYVLVDHAIFFRAENKGAALILKYEGEENHIEYHKQDAVYL	369		
Db	305 IVPAGGSAIVEFPKVDIPGSYTLVDHSHFRANKGALGGLKVEGAENPEIMTOKLSDTAYA	364		
QY	370 PEGAPQAIIDTQEAKTAPAPA 389			
Db	365 GNGAAPA---ASAPAPASAPA 381			
RESULT 3				
Q9JTB8	PRELIMINARY;	PRT;	386 AA.	
ID	Q9JTB8;			
AC	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-OCT-2002 (TREMBlrel. 22, Last annotation update)			
DE	Nitrite reductase, major outer membrane copper-containing protein (EC 1.7.99.3).			
GN	AN1A OR NMA1887.			
OS	Neisseria meningitidis (serogroup A).			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;			
OC	Neisseriaceae; Neisseria.			
OX	NCBI_TaxId=65699;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=Z2491 / Serogroup A / Serotype 4A;			
RX	MEDLINE=20222556; PubMed=10761919;			
RA	Parkhill J., Achman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrrell B.G.;			
RA	"Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."			
RL	Nature 404:502-506(2000).			
DR	EMBL; AL162757; CAB85110.1; .			
DR	HSSP; P38501; 1AS8.			
DR	InterPro; IPR001117; Cu-oxidase.			
DR	InterPro; IPR001287; CUNO2_reductase.			
DR	Pfam; PF00394; Cu-oxidase; 2.			
DR	PRINTS; PR00695; CUNO2RDTASE.			
DR	Oxidoreductase; Complete proteome.			
QW	SEQUENCE 386 AA; 40389 MW; D0624ED3A3979E88 CRC64;			
Query Match	52.8%;	Score 1386;	DB 16;	Length 386;
Best Local Similarity	68.7%;	Pred. No. 1.4e-85;		

Query Match	52.8%;	Score 1386;	DB 16;	Length 386;
Best Local Similarity	68.7%;	Pred. No. 1.4e-85;		

Matches	263;	Conservative	45;	Mismatches	69;	Indels	6;	Gaps	2
Qy	7	IKTTLICALSALMISGCSNQADKAQPKSSTVDAAKTAKNADNAPASQEHQGEIPVDAIV	66						
Db	1	MKQQLAALIAFMFLAACGGEPAAQTPTAASAAASA--QTAETAGELPVDAVT	57						
Qy	67	THAPEVPVDDSDPAKVVVKMETEYKVRLLADGVEYQWTFGGVPGQMIRVEDGTE	126						
Db	58	THAPEVPAIDRDYPAKRVVKMETEYTKMDDGVEYRYWTDGDVPGMIRVEDGTE	117						
Qy	127	VQPSNHPDSKMPBNVDFFAATPGGGAASFTAPGHTSTFSFKALOPGLVYVCAAPVG	186						
Db	118	VBFSSNPSTVPHNVDFFAATQGGGAALTFAFGTSTFSFKALOPGLIYHCAVAPVG	177						
Qy	187	MHIANGMYGLILVEPREGI.PKYDKKEYVWQDFFYTKGYGEOGLQPFDEKAIREDAEVY	246						
Db	178	MHIANGMYGLILVEPREGI.PKYDKKEYYIYQDGFYTKGKGAGAGLOPFDMDKAIAREPEVY	237						
Qy	247	VFNQSGVALTGENALKAKVGETVRL.FVNGGNGNLTSFVHIGEIFDKVAREGGKGENHI	306						
Db	238	VFNHGVGAIAGDNALKA.KAKGETVRYMYVNGGNGNLVSSFVHIGEIFDKVYVEBGGKLINEV	297						
Qy	307	QTTLIPAGCAAITERRVDVPGDYVLVDHAI.FFAFNKGAIGILKVEGENHEIYSHKQIDA	366						
Db	298	OSTIVPAGGSAIVEKVDIPGSYTLVDHSH.FFAFNKGAIGQLKVEGAENPEIMTQGLSQT	357						
Qy	367	VYLPBGAPQALDTCAPKTPAPA	389						
Db	358	AVAGNGAAPA--ASAPPAASA	377						

ID	Q8KXH4	PRELIMINARY;	PRT;	486 AA.
AC	Q8KXH4;			
DT	01-OCT-2002 (TREMBLrel. 22, Created)			
DT	01-OCT-2002 (TREMBLrel. 22, last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, last annotation update)			
DE	Nitrite reductase precursor.			
GN	NIRK.			
OS	Hyphomicrobium denitrificans.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
CC	Hyphomicrobiaceae; Hyphomicrobium.			
OX	NCBI_Taxid=53399;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3151;			
RA	Kataoka K., Fukui A., Kabayashi M., Yamaguchi K., Suzuki S.;			
RT	"Cloning and expression of copper-containing nitrite reductase from			
RT	Hyphomicrobium denitrificans."			
RL	Submitted (DEC-2001) to the EMBL/genbank/DBJ databases.			
DR	EMBL; AB076606; BAC00912.1; -			
DR	InterPro; IPR001177; Cu-oxidase.			
DR	InterPro; IPR001287; CuNO2 reductase.			
DR	PRINTS; PR00695; CUNO2RDTASE.			
DR	PROSITE; PS00079; MULTICOOPER_OXIDASE1; 1.			
KW	Signal.			
FT	SIGNAL.			
SO	SEQUENCE	1	39	
		486 AA;	52238 MW;	6984E02E6CF791AF CRC64;
Query Match		30.0%;	Score 787;	DB 2; Length 486;
Best Local Similarity		51.1%;	Pred. No. 5,6e-45;	
Matches 157;	Conservative 41;	Mismatches 105;	Indels 4;	Gaps 3
QY	55 HQGELPVDALVTNAP-EVPPVDDHDHAKVYVVKMEYKVKRLADGVGEQFWTFGGQVP	113		
Db	173 NRAEMKSSGADITRPADLPDPIGPQAKTVRIDETLYEVKQQLDDNTTYITWTNGKVP	232		
QY	114 GQMIRVREGDITLEVQFSNHPDSKMHANDFNAATGPGGGAELSFAPGHTSTFSFKALDP	173		
Db	233 GPFLEVRGDTVELHLKNHKOSLMHNSVDHGATGPGGAAATQDDPGEEIVTFKALIP	292		
QY	174 GLVYVHCVAAPVGMHIAAGMYGLIIVERKEGIPKVDKEVYVWQDGFYTGKYGEOGLQPF	233		

Db	23	GIVVHCATPSVPHITITNGMYELLIVERPGGLPQVDRREYVNGGELIYTTKSPFITSGEQM	352
QY	234	DMEKAIREDAEVFNNGSVGALTGENALKAKVGETVRLTFVNGCGPNLTSSFHVIGEIFDK	239
Db	353	DYEKLINEPEEFELFNNGSVGLTRSHPLTASVGETVRIFFVGCGPNPTSSFHVIGEIFDH	412
QY	254	VHEFGG-KGEHNHQTTLIPAGGAITFEKVDVPGDYVLVDHAIIFRAFNKAGLGIKVE	351
Db	413	VYSLGSVSPPLIGVQTVTSVPPGCGATVDFKIDRAGRITLLVDHALSR-LEHGLVGFPLNV	471
QY	352	GEENHEI	358
Db	472	GEKNDSI	478

RESULT 5			
ID	Q9P9H9	PRELIMINARY;	PRT; 361 AA.
AC	Q9P9H9;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)		
DT	01-OCT-2002 (TREMBLrel. 22, last annotation update)		
DE	Copper-containing dissimilatory nitrite reductase precursor (Ec 1.7.99.3).		
GN	NIRK.		
OS	Halobaculum marismortui (Halobacterium marismortui).		
OC	Archaea; Euryarchaeota; Halobacteriota; Halobacteriales;		
OC	Halobacteriaceae; Halobaculum.		
OX	NCBI_TaxID=2238;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21311739; PubMed=11418554;		
RA	Ichiki H., Tanaka Y., Mochizuki K., Yoshimatsu K., Sakurai T.,		
RA	Fujiwara T.,		
RT	"Purification, characterization, and genetic analysis of Cu-containing		
RT	dissimilatory nitrite reductase from a denitrifying halophilic		
RT	archaeon, Halobaculum marismortui";		
RL	J. Bacteriol. 183:4149-4156(2001).		
DR	EMBL; AJ278286; CAB93142.1; -.		
DR	HSSP; P38501.1a88		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR001287; CuNO2_reductase.		
DR	Pfam; PF00394; Cu-oxidase; 1.		
DR	PRINTS; PR00695; CUNO2RDTASE.		
KW	Oxidoreductase; Signal.		
FT	SIGNAL 1 31		POTENTIAL.
FT	CHAIN 32 361		COPPER-CONTAINING DISSIMILATORY NITRITE
FT			REDUCTASE.
SEQUENCE	361 AA; 38803 MW; 12829C586AA3F5D0 CRC64;		

Query Match	28.6%	Score 751.5	DB 1	Length 361
Best Local Similarity	42.5%	Pred. No. 9,2e-43		
Matches 154	Conservative 55	Mismatches 138	Indels 15	Gaps 5
QY	PTLIITTLICL--SALMLSGSNQADRAQKSKSTVDAATTAATNANDNAASCHQSELPV	61		
DB	5 PTATRRRLALEALGVGTALALAC-----ASAPCAAEQATEATTTPDEPMNNAAQDND---	55		
QY	IDAIVTAHAEVPPRPVDRDHPAKVYVYKMETVEKVMRLADGEVQYFWFGGQVPGOMTRVRE	121		
DB	56 VDRIDADPTALPDPIDRSEPKTVSEVMTTKQVLAIEBEGVTTYTMFGQOIEPQPMTRVRR	115		
QY	122 GDTEIVQFSNHPDSKMPHNVDFHAATGPGGAGASFTABGHTSTGSKFALQPLVYVNHCA	181		
DB	116 GDTEVLTITTEEGNSMPPNIDILHAVRGCGGGAELAMVTEPGQKTFEPFKATVPCGAFIYNCA	175		
QY	182 VAPQOMHTANGMYGLILVEPEKGLPKVDKEYVVMQDFTYKSKYEGQGLQPEMEKAIKE	241		
DB	176 VPNIDMTLSSGPMGMITLVEPKEGLEVDHEFEGQHELTITTDGTGKGHHDDPEMEMAAB	235		
QY	242 DAEVYVFGSGVQALTGE--NALKAKVGETVRLFYVNGGNGPNLTSSPFVIGEIFDKVYFEGG	299		

Db 226 EPTVLIAMNGEKVAITPDRGSPQMVGETARVYFVTGGPRLDSSFPHISGMBEVMQGS 255

QY 300 KG - BNHNQTTLLIPAGGAITEFEKVDVGDVYLDYHAIIPRANKALGLIKVGEENHE 357

Db 226 IAGPFRNRYQTTFPVKGSCALATLHAEVGFPIKLDYHALLSRVARKKTMALINNEGAAAPD 355

QY 358 IY 359

Db 356 VF 357

RESULT 6

ID	Q9AC07	PRELIMINARY;	PRT;	360 AA.
AD	Q9AC07			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Disimilatory nitrite reductase.			
GN	NIRK.			
OS	Alcaligenes sp. STC1.			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC	Alcaligenaceae; Alcaligenes.			
OX	NCBI_taxid=133923;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=STC1;			
RA	Shoun H., Takaya N.;			
RT	"Alcaligenes sp. STC1 nitrite reductase gene."			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB046603; EMBR2510.1; -.			
DR	HSSP; P25006; INTF.			
DR	InterPro; IPR001117; Cu-oxidase.			
DR	InterPro; IPR001287; CuNO2 reductase.			
DR	InterPro; IPR002160; Kunitz legume.			
DR	Pfam; PF00394; Cu-oxidase; 2.			
DR	PRINTS; PR00695; CUNO2REDTASE.			
DR	PROSITE; PS00283; SOYBEAN_KUNITZ; 1.			
SQ	SEQUENCE 360 AA; 39072 MW; CAF063CB38430380 CRC64;			

RESULT 7

```

068601
ID 068601 PRELIMINARY; PRT; 360 AA.
AC 068601;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Desimilatory copper-containing nitrite reductase.
GN
OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Achromobacter.
CX NCBI_TaxID=515;
[1]
RN
RP SEQUENCE FROM N.A.
RP STRAIN=NCIMB11015;
RA Suzuki E., Horikoshi N., Kohzuma T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=GIFU051;
RA Katsuka K., Furusawa H., Yamaguchi K., Suzuki S.;
RT "Cloning and Expression of Copper Nitrite Reductase Gene from
RT Alcaligenes xylosoxydans GIFU051."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
EMBL AF051831; AAC05831.1; -
EMBL AB013078; BAA33678.1; -
DR HSSP; P25006; INTF.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR001287; CUNO2_reductase.
DR Pfam; PF00394; Cu-oxidase; 2.
DR PRINTS; PR00695; CUNO2RDTASE.
KW Signal
SQ SEQUENCE 360 AA; 38939 MW; 374085BD3BFF44E7 CRC64;

```

RESULT

Q8U7R7	
ID Q8U7R7	PRELIMINARY; PRT; 455 AA.
AC Q8U7R7;	
DT 01-JUN-2002 (Tremblay, 21, Created)	

Db 372 APBG 375

RESULT 10

08YB82 PRELIMINARY; PRT; 376 AA.

AC 08YB82;

DT 01-MAR-2002 (TEMBLrel. 20, Created)

DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)

DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Copper-containing nitrite reductase precursor (EC 1.7.99.3).

GN BMEI10988.

OS *Brucella melitensis*.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI_TaxId=29459;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=16M / ATCC 23456 / Biotype 1;

RX MEDLINE=20020109; PubMed=11756688;

RA Delvecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T., Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G., Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E., Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J., Haselkorn R., Kyriides N., Overbeek R.;

RA "The genome sequence of the facultative intracellular pathogen *Brucella melitensis*."

RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).

RL EMBL; AF009732; AAL54230.1; -.

DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR001287; Cuno2_reductase.

DR InterPro; IPR006311; Tac.

DR Pfam; PF00394; Cu-oxidase; 2.

DR PRINTS; PR00695; CUNO2RDTASE.

DR TIGRFA; TIGR01409; TAT_signal_seq; 1.

DR Oxidoreductase; Complete proteome.

DR KW

SQ SEQUENCE 376 AA; 40515 MW; 4D95E3F21C105AF9 CRC64;

Query Match 16.4%; Score 431; DB 16; Length 376;

Best Local Similarity 34.8%; Pred. No. 4.3e-21;

Matches 122; Conservative 48; Mismatches 147; Indels 34; Gaps 11;

Db 43 KTANADNAAQOHEGELPVDAITVTHAEPVPRVDHRAKVVVKMETEKKWRLAD-GV 101

Db 38 RKASAEIALPRQ-KVELDPRPVHA--HTQVAEGRPVVOFTVIEKKIVIDDAGT 93

QY 102 EYQWTFGGVPGQOMIRVREGDTIEVOFSNHPDSKMPHNVDFHAATGPGGAASFTAPG 161

Db 94 EVHAMTFNGTVPRPLMVHDDYELTLINPENTNLLHIDFHAATGALGGGLTEINPG 153

QY 162 HTSTSFKALQPLLYTHCAVAPVGM--HIANGMGLIVERKEGLP-----KYDKE 211

Db 154 EKTVLRFKATKPCVFVYHC--APPGMVPMHVIVSGMGAANVLLPREGLHDKGNKLTVDKY 211

QY 212 YVMOGDFYTK-----GKYGEQGLQPFMEKAIRE-DAEYVFNQSGVGLTGENLKA 263

Db 212 YVIGEDDFYPRBENGNYKTYEARGDVETVKMRLTLTTHVFNCAVALGDRALTA 271

QY 264 KVEETVALFVNGGPNLTSSFHVIGELFDKVNHEGKGENHNT--QTLTLPAGGAITEP 321

Db 272 KVEEKLVI--HSGQANDTRPHLIGCHGDYVMAATGKENTPRPDVTFPGAAAAAFY 329

QY 322 KVDVPGDYVLVDHAIFAFPAFKGALGLKVEGENHEIYSHKQDVAIVLPBG 372

Db 330 TFRPGIYAVVNHNLTEAFELGAAPKVGEMNDLMT-----SLAPSG 375

RESULT 11

031380 PRELIMINARY; PRT; 364 AA.

AC 031380;

DT 01-JAN-1998 (TEMBLrel. 05, Created)

DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)

DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)

DE Respiratory nitrite reductase.

GN NIKK.

OS *Bradyrhizobium japonicum*.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bradyrhizobiaceae; Bradyrhizobium.

OX NCBI_TaxId=375;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=USD110;

RA Velasco L., Bedmar E.J.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ002516; CA00506.1; -.

DR HSSP; P38501; 2AFN.

DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR001287; Cuno2_reductase.

DR Pfam; PF00394; Cu-oxidase; 2.

DR PRINTS; PR00695; CUNO2RDTASE.

SQ SEQUENCE 364 AA; 39356 MW; D6C197968F290714 CRC64;

Query Match 16.4%; Score 429.5; DB 2; Length 364;

Best Local Similarity 33.4%; Pred. No. 5.2e-21;

Matches 126; Conservative 50; Mismatches 146; Indels 55; Gaps 12;

Db 14 ALSALMLSGSCNQADKAAPKSSITVDAAKTANADNAAQOHEGELPVDAITVTHAEPV 73

Db 12 AATLMTLAPRALADDLKPRKVELVAPRFVHAHQATK--QG----- 53

QY 74 PVVDHRAKVVVKMETVE--KVMRLADGEYQFTFGGQVPGQOMIRVREGDTIEVOFSN 131

Db 54 -----PKIVPEKLTIEKKVVIDEKGTFQAMTFNMGPMPLMVHGGDYVETTLVN 105

QY 132 HPDSKMPHNVDFHAATGPGGAASFTAPRGHSTSFKALQPLLYTHCAVAPVGM--M 187

Db 106 PATNTPHNIDPHSATGALGGALTLINDEGVVLRWAKTKGVYHC--APGPMIPW 163

QY 188 HIANGMGLIVERKEGLP-----KYDKEYVMQGFYT---KG---KYGEQGLQPF 233

Db 164 HVSQMGNGAVWVLPBGLNDGKHALKYDKVYVVGSDMVVPRDEKGNFKSYDSFGAEST 223

QY 234 DMEKAIRE-DAEYVFNQSGVGLTGENLKAAYGETVRLFVNGGPNLTSSFHVIGELFD 292

Db 224 DTEEMWKKLIPSHVFNQSGVGLTGENLTAANVENV--LIHSGQANDSRPHLIGHGD 281

QY 293 KYHFEQKGENH--NIQTLTLPAGGAITEPKVVDVPGDYVLVDHAIFAFPAFKGALGLK 349

Db 282 YV-WETGKGNAPDEVGLETWFTFGSAGAMTKFMQGIYAVVTHNLTEADLCATAPK 340

QY 350 VEGEENHEIYSHKQDA 366

Db 341 VEGKWNDDLMTQVAPA 357

RESULT 12

08FX29 PRELIMINARY; PRT; 376 AA.

AC 08FX29;

DT 01-MAR-2003 (TEMBLrel. 23, Created)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Copper-containing nitrite reductase.

GN BRA0260.

OS *Brucella suis*.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI_TaxId=29461;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=1330 / Biovar 1;

RX MEDLINE=22247741; PubMed=12271122;

RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F., Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J., Daugherty S.C., Debay R.T., Durkin A.S., Kolonay J.F., Madupu R.,

RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettele H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.,
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts";
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL: AB014526; AAN33462.1; -.
KW Complete proteome.
SQ SEQUENCE 376 AA; 40499 MW; 4C0CF7A45FAB1BEA CRC64;

Query Match 16.3%; Score 429; DB 16; Length 376;
Best Local Similarity 34.5%; Pred. No. 5.9e-21;
Matches 121; Conservative 49; Mismatches 147; Indels 34; Gaps 11;

QY 43 KTNADNNAASQEHQGLPVITDAIVTHAPVPPVDDHAPKVVVKETEKWRLAD-GV 101
DB 38 RQASAEIDALPRQ-KVELVDPEFVHA--HVOVAGSGKVVQFTVIEKKIVIDAGT 93
QY 102 EYQFTWFGQVPPQGMIRREGDTIEVQFSNHPDSKPHVDPHATGPGGGAASFTAPG 161
DB 94 EYHMTFNGTVPGLVAVVHQQDDYLETLINPETNTLLHNI DPHATGALGGGLTEINPG 153
QY 162 HTSTSFRLQRLGLVYVHCAVAVPGM--HIANGTGLLVPEKGLP-----KVQKE 211
DB 154 EKTVALRFKATKPGVFVYHC--APPGVPMVHVSGMNGAVVVPREGLDHDKNKLTYYDKV 211
QY 212 YVVMGDDFYTK-----GKYGEQGLQPFDMEKALRE-DAEYVFNQSVGALTGENALKA 263
DB 212 YVVGQDFVPPDENQNTYTPAPGDAYEDYVAKWMTLTPTTHVENGAGALITGDALTA 271
QY 264 KVGETVRLVFGNGGPNLTSSFVIGEIFDKVHEGSGKGNHNI--QTLTIPGGAITFEF 321
DB 272 KSGEKVLIIT--HSQANRDRPHLIGHGDIYVATGKFNTPPVDDGTWTFPGAAAAAFY 329
QY 332 KYDVGDDVYLVDHAI FRAFNKAGALGLKTEGSENHETYSHKQTDVATVLEPG 372
DB 330 TFRQGIYAVVYVHNLIEAFELGAVHAFKVTGEMNDLMT-----SLAPSG 375

RESULT 13

QY 08FTV6 PRELIMINARY; PRT; 478 AA.
AC 08FTV6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative major outer membrane protein.
GN CE0272.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxId=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.,
RT "The entire genomic sequence of Corynebacterium efficiens YS-314";
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AP005214; BAC17082.1; -.
KW Complete proteome.
SQ SEQUENCE 478 AA; 50578 MW; 282BD1715C02D4A CRC64;

Query Match 14.9%; Score 390.5; DB 16; Length 478;
Best Local Similarity 31.4%; Pred. No. 3.3e-18;
Matches 108; Conservative 47; Mismatches 130; Indels 59; Gaps 13;

QY 31 AOPKSTVDAAKTANADNAASQEHQGLPVITDAIVTHAPVPPVDDHAPKVVVKET 90
DB 168 AAPAGCTPAADVAVPGV-GARSVDDHGGFTVPVLT-----APAGTVH-----YSMDI 216

QY 91 VEKWRDLADGVYQFTWFGQVPPQGMIRREGDTIEVQFSNHPDSKPHVDPHAA-TCP 149
DB 217 TEVEVQVAPGHQVAVLDFGQAPPTLGRGLDTRITL--HNKGTMDHSIDFHAGEVNP 274
QY 150 GGAASFTAPRPHSTSTFESKALQPLGLVYVHCAVAVGHNHIANGTGLLVE-----KEEL 205
DB 275 --DTNMAQTPVGESLTIEFVANRYGIMWHCTAFTMSLHIANMGCAVITIDPADSAESL 332
QY 206 PKYDEEYVMOGDFYTKGKGQGLQPFDMEKALREDAEYVFNQSVGALTGEN----- 259
DB 333 SEVDEEYLVASSEMFG-----ETELGADQAV--NDRDYDLTAPNFPNRY 377
QY 260 ---ALKAVGETVRLVFGNGGPNLTSSFVIGEIFDKVHEG-----GKGNHNI 306
DB 378 DLAPIEHKVGDTVRVWLNKVGPDQSLSPHVGEVDFTVFSEGRYILIRDASTRGTGS---- 433
QY 307 QTLTPAGGAATTEKVDVPGDYVLVDHAI FRAFNKAGALGLKY 350
DB 434 QAVNVSAAGAFIELTFNAPGNVAFVNHQMTDA-BKGHGFPYV 476

RESULT 14

QY 08FSM1 PRELIMINARY; PRT; 343 AA.
AC 08FSM1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN LA1555.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxId=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB011334; AAN48754.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 343 AA; 38017 MW; 6D850CA0F12DB9D3 CRC64;

Query Match 9.8%; Score 258; DB 16; Length 343;
Best Local Similarity 28.2%; Pred. No. 1.8e-09;
Matches 78; Conservative 41; Mismatches 120; Indels 38; Gaps 11;

QY 75 PVDHDPKAVVVKETVEKWRDLADGVYQFTWFGQVPPQGMIRREGDTIEVQFSNHPD 134
DB 95 PILKH-----ISIVEMPLVANHNTVAKMTFGLVPGVPRAKLGQRMETLRN--D 145
QY 135 SKMPHNVDPHATGPG-GGAASFTAPRPHSTSTFESKALQPLGLVYVHCAVAVGHNHIANGM 193
DB 146 SEHPSHIFHSGHDNEDGWER--VVQSEKTYQTLAPRIGHNPHCHVPPPLASHMAGSL 203
QY 194 YGLIVPEKEGLPKYDEEYVMOG-DEYTKGKGQGLQPFDMEKALREDAEYVFNQSV 252
DB 204 YGGLVDPGRGRPRNHEMMLISGDLDEDKK-----NDLFGMNGMA 245
QY 253 GALTGENALKAVGETVRLVFGNGGPNLTSSFVIGEIFDKVHEGSGKGNHNIQTLT 311
DB 246 G-FYDRIYIKVPVQGVKLYTIANMCEYRPVASFHLHAQTFFV--FRTGTLKLRDHTTV 302
QY 312 PAGGA--ATTEPKVDVPGDYVLVDHAI FRAFNKAGLG 346
DB 303 TLGQTERVILEFTLTPKRGYMHFHPQTKMA-EKGAMG 338

RESULT 15

QY 093TN1 PRELIMINARY; PRT; 179 AA.
AC 093TN1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
GN NIKR.
OS Nitrosomonas sp. TA-9211-NH4.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=159355;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TA-9211-NH4;
RX MEDLINE=21218534; PubMed=11319103;
RA Casagioti K.L., Ward B.B.;
RT "dissimilatory nitrite reductase genes from autotrophic ammonia-
oxidizing bacteria";
RL Appl. Environ. Microbiol. 67:2213-2221 (2001).
DR EMBL; AF339049; AK53363.1;
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1 179
FT SEQUENCE 179 AA; 20005 MW; 85AOCFEF7963B12D CRC64;
SQ
Query Match 9.6%; Score 253; DB 2; Length 179;
Best Local Similarity 39.6%; Pred. No. 1.6e-09;
Matches 74; Conservative 23; Mismatches 58; Indels 32; Gaps 11;
QY 177 VYHCANAVPVGMM--HANGVYGLIVPEKGLPK-----VDKEYVMQGDFFY----TK 222
DB 1 VYHC--APGGMVPMHVGSMGTIMVLPKGLKPKGLHYDRIVYIGENDYIPKDN 58
QY 223 GKX-----GEOGLQFPD--MEKAIREDAYVFNVSAGALTGENAKAKVGETVRLFVGN 275
DB 59 GKXKYVDSPEAPADTTDIRKLI---PNHVFNGRVGALTGNKATKAGETV-MFV-H 113
QY 276 GGFNLTSFHVIGEIFDKVHFEGSGKEN--HNIOITLLIPAGGAITEFKVDVPGDYVLV 332
DB 114 SEMNRDTRPHIIGLDGYV-WETGKFSNPEKPLETFIFIGSAGALYFRQPGVYAVY 172
QY 333 DHAIFRA 339
DB 173 NHNLIEA 179
RESULT 16
QY0904 PRELIMINARY; PRT; 379 AA.
ID Q9HOF4
AC Q9HOF4; (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
DE Membrane protein.
GN PANI OR VNG1187G.
DN Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20504483; PubMed=11016950;
RX NG W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lesky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Meli R., Goo Y.A.,
RA Leitauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonki P.E., Kreb M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlischer M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasgupta S.;
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
DR EMBL; AE005046; AAC19561.1;
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 2.

DR TIGRfam; TIGR01409; Tat_signal_seq; 1.
KW Complete proteome.
SQ SEQUENCE 379 AA; 41173 MW; 4807DF07F8AA666 CRC64;
Query Match 9.5%; Score 250.5; DB 17; Length 379;
Best Local Similarity 25.3%; Pred. No. 6.8e-09;
Matches 94; Conservative 52; Mismatches 133; Indels 93; Gaps 20;
QY 20 LSGCNSQADKAAQPKSSIVDAAK-----TANADNAASQHGELVDAIVTHAREVPP 75
DB 29 IAGCT--APDGESEVDTTAKKQGLPTSPPEVDAITEGQNVTLKSV-----PA 78
QY 76 VDRDPAKVV--VNETVEKVMRLADGVEYQFTFGGQ----VPGMIRVREGDTIEV 127
DB 79 VHDVPLDSMGSPVYLPV-----WAFATEDGSPVPGPIVRETEGQDLE 124
QY 128 QFSNHPDSKMPHNVDFHAATGPGGAENASF-----TANGHTSTSFKALQGL 175
DB 125 TLDN--TDGKRPHTLHFH-----GSQTAMEDDGVPTTGIRVGEKRTYTI PANVPQT 176
QY 176 VYHCANAVPVGMMHANGVYGLIVPEKGLPKVDKEYVMQGDFFTKGKYGGLQFPDM 235
DB 177 HLYHGY-OTQRIIDMGMYGIFRIDPK-GYEPADKYEFTVTKWDMSRLNSHAG----- 228
QY 236 EKAIREDAEYV--VENGSGVALLTGENA-----LKAKVGETVRLFVNGGPNLTS 282
DB 229 -----EDVDYSPRTNPNPVFTVNGKSAPRTLHPEDGSPVIVGSDTVRLHLVNGG-YMH 282
QY 283 SFHVIGEIFDKVHFEGK--GENNIO-TLLIPAGGAITEFKVDV-PGDYVL----VD 333
DB 283 PLHINHFPRVKEKGGQIPAAQHTMTDITMAPARHTI-EFOADADPGIYIMCHKVN 341
QY 334 HAIFRAFNKAL 345
DB 342 HVNMGTFYPGM 353
RESULT 17
ID Q8YO92
AC Q8YO92; PRELIMINARY; PRT; 338 AA.
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
DE Hypothetical protein A113942.
GN ALU3942.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21595285; PubMed=11759840;
RX Kaneko T., Nakamura Y., Wolk C.P., Kurihara T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AP003594; BAB75641.1;
DR InterPro; IPR001127; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
DR PRINTS; PR00695; CUNO2_REDUTASE.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 338 AA; 38408 MW; AFB1B9D01627F7 CRC64;
Query Match 9.5%; Score 248.5; DB 16; Length 338;
Best Local Similarity 27.8%; Pred. No. 7.8e-09;
Matches 76; Conservative 42; Mismatches 116; Indels 39; Gaps 12;
QY 94 VNRADGVEYQFTFGGQVPGQMIRVREGDTIEVQFSNHPDSKMPHNVDFHAATGPGGA 153

Db 93 VQOLNSAVSNINWDLNGRIIPGTLRKQKGRIRVLP--HNQASHSLHFHGV----HRA 146
154 EASFTAP---GHTSTSPKALQPLGVYVHCAVAPVGMHTANGVGLIVEPKGGLPKVX 210
147 EMDGVRPVNNKSTIYEPDAPRYGVHLTHCHIEPVRHAKSLYGMFTIDPPRPAPD 206
Qy 211 EYVWQGDFTYTKGKGEGGLQPPDMEKAIREDAYVFNVSAGALTGENALKAKVETVA 270
Db 207 IYLVNMG-----YVDVDSHND--FAFNGLPNHNW-DNPICQYONQILR 248
Qy 271 LEVGN---GGPNTSSFVHIGIFDKVHVEGKGNNHNTQTLIPAGAA--ITERKVDV 325
Db 249 LYVNLINIEYDPAVT--FHLHAFEDVYRY--GSMKASEKTIDVITGVARHILEPAPRY 304
Qy 326 PGDYVLVDHAIFFAFKAGALILKY--EGEENH 356
Db 305 PGKYMFPHQDAIAEN-GCMGQFEVAVANNONH 336

RESULT 18

OBRUT7

ID OBRUT7; PRELIMINARY; PRT; 179 AA.

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
GN NIRK.
OS Mesorhizobium sp. 4FB11.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
RN NCBI_TaxID=126151;
RP SEQUENCE FROM N.A.
RC STRAIN=4FB11;
RA Song B., Ward B.B.;
RT "Nitrite reductase genes in halobenzate degrading denitrifying
bacteria."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY078254; AAL82513.1; -
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT 179
SQ SEQUENCE 179 AA; 19893 MW; 0060433CAD0A516C CRC64;

Query Match 9.0%; Score 236; DB 2; Length 179;
Best Local Similarity 37.8%; Pred. No. 2.2e-08;
Matches 70; Conservative 27; Mismatches 60; Indels 28; Gaps 10;

Qy 177 VVHCAVAPVGM---HIANGMGLILVEPKGP-----KYDKYYWQSGPYTKGKYG 226
Db 1 VVHC--APGGMTPMVVSGMHTIVLPPDGLKNEKGLVKDITYYIDNNEFYIP-RDE 57
Qy 227 EGGLOPFD-MEKAIREDAE-----YVFNVSAGALTGENALKAKVETVRLFPVNGG 277
Db 58 DGNFKKPSLSGESYETMEVWGLIPTHVFNKRGSLTDNAMSKEVETV-LFY-HSQ 115
Qy 278 PNLTSFVHIGEIFDKVHVEGKGNNHNTQTLIPAGAAITEFKVDVPGDYVLVDH 334
Db 116 ANRDRPRLHIGHGIDYV--WEGKFNAPRAKDLFTWIFRSGSAGALYTLQGVAVVNH 174
Qy 335 AIFRA 339
Db 175 NLIEA 179

RESULT 19

OBRUT4

ID OBRUT4; PRELIMINARY; PRT; 179 AA.

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
GN NIRK.
OS Ensifer sp. 2FB8.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Ensifer.
RN NCBI_TaxID=126147;
RP SEQUENCE FROM N.A.
RC STRAIN=2FB8;
RA Song B., Ward B.B.;
RT "Nitrite reductase genes in halobenzate degrading denitrifying
bacteria."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY078247; AAL82506.1; -
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT 179
SQ SEQUENCE 179 AA; 19695 MW; D94A75486978BD7A CRC64;

Query Match 8.9%; Score 233; DB 2; Length 179;
Best Local Similarity 38.6%; Pred. No. 3.5e-08;
Matches 71; Conservative 22; Mismatches 65; Indels 26; Gaps 8;

Qy 177 VVHCAVAPVGM---HIANGMGLILVEPKGP-----KYDKYYWQSGPYTK----- 222
Db 1 VVHC--APGGMTPMVVSGMHTIVLPPDGLKDEKGPITYDKIYYGEODFYVPKDEA 58
Qy 223 ---GKYGEGLQPPDMEKAIREDAYVFNVSAGALTGENALKAKVETVRLFPVNGG 278
Db 59 GNYKYEETPGEAYEDAVAMRLTPTHVFNKRGSLTDNAMSKEVETV-LFYVHQA 116
Qy 279 NLTSSFVHIGEIFDKVHVEGKGNNHNTQTLIPAGAAITEFKVDVPGDYVLVDH 335
Db 117 NNDTRPRLHIGHGIDYV-WATGKFRNPDLDETWLIPGTAGAAFYTFPQGVAVVNH 175
Qy 336 IFRA 339
Db 176 LIEA 179

RESULT 20

OBRUT3

ID OBRUT3; PRELIMINARY; PRT; 179 AA.

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
GN NIRK.
OS Ensifer sp. 4FB6.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Ensifer.
RN NCBI_TaxID=126148;
RP SEQUENCE FROM N.A.
RC STRAIN=4FB6;
RA Song B., Ward B.B.;
RT "Nitrite reductase genes in halobenzate degrading denitrifying
bacteria."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY078248; AAL82507.1; -
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT 179
SQ SEQUENCE 179 AA; 19594 MW; B9EA63390B6CF6E CRC64;

Query Match 8.8%; Score 232; DB 2; Length 179;
Best Local Similarity 38.2%; Pred. No. 4.1e-08;
Matches 71; Conservative 20; Mismatches 65; Indels 30; Gaps 8;

```

QY 177 VYHCAVAPVGM---HIANGMVGLIIVEPKKGLP-----KVDKEYVMQGDIFYTK----- 222
DB 1 VYHC--APGGMVPMWHTSGMNGAIWVLPBDGLKDEKQGPLTYDKIYVGEOPFYPKDEA 58
QY 223 ---GKYGEQGLQPFDMKEAIRD---DAEYVFNKSVGALGTENALKAKEVETVRLFVNGG 278
DB 59 GNYKKTETGEAYEDVAKMRITLPTHTYFNGVVGALTDHDLTAAGV--RVLVHSA 116
QY 279 NLTSFHHVIGEIFDKVHFEKGKGNHN-----IQTLIPAGGAATTEFKVDVPGDYVLVD 333
DB 117 NADTRPHLIGHGDYV---GATGKFPNPPDLQDETWLIPGTAGAFYTRQPGVAYAVN 173
QY 334 AIFRA 339
DB 174 NMLEA 179

RESULT 21
ID 093TN6 PRELIMINARY; PRT; 180 AA.
AC 093TN6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
OS Nitrosomonas marina.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_Taxid=917;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-56;
RC MEDLINE=21218534; PubMed=11319103;
RA Casciotti K.L., Ward B.B.;
RT "Dissimilatory nitrite reductase genes from autotrophic ammonia-
oxidizing bacteria."
RL Appl. Environ. Microbiol. 67:2213-2221 (2001).
DR EMBL; AF339044; AAK5358.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT 180
SQ SEQUENCE 180 AA; 19681 MW; 43BBD2BE098968FE CRC64;

Query Match
Best Local Similarity 34.1%; Pred. No. 5.2e-08;
Matches 63; Conservative 33; Mismatches 62; Indels 27; Gaps 9;

QY 177 VYHCAVAPVGM---HIANGMVGLIIVEPKKGLP-----KVDKEYVMQGDIFYTK-----T 221
DB 1 VYHC--APGGMVPMWHTSGMNGAIWVLPBDGLKDEKQGPLTYDKIYVGEOPFYPKDEA 58
QY 223 ---GKYGEQGLQPFDMKEAIRD---DAEYVFNKSVGALGTENALKAKEVETVRLFVNGG 277
DB 59 GNYKKTETGEAYEDVAKMRITLPTHTYFNGVVGALTDHDLTAAGV--RVLVHSA 116
QY 279 NLTSFHHVIGEIFDKVHFEKGKGNHN-----IQTLIPAGGAATTEFKVDVPGDYVLVD 334
DB 117 NADTRPHLIGHGDYV---GATGKFPNPPDLQDETWLIPGTAGAFYTRQPGVAYAVN 173
QY 334 AIFRA 339
DB 176 NMLEA 180

RESULT 22
ID 093TN3 PRELIMINARY; PRT; 180 AA.
AC 093TN3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

```

```

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
OS Nitrosomonas sp. C-45.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_Taxid=159353;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-45;
RC MEDLINE=21218534; PubMed=11319103;
RA Casciotti K.L., Ward B.B.;
RT "Dissimilatory nitrite reductase genes from autotrophic ammonia-
oxidizing bacteria."
RL Appl. Environ. Microbiol. 67:2213-2221 (2001).
DR EMBL; AF339047; AAK53361.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT 180
SQ SEQUENCE 180 AA; 19540 MW; BB7CA048F7CD6D29 CRC64;

Query Match
Best Local Similarity 35.1%; Pred. No. 7.1e-08;
Matches 65; Conservative 30; Mismatches 63; Indels 27; Gaps 9;

QY 177 VYHCAVAPVGM---HIANGMVGLIIVEPKKGLP-----KVDKEYVMQGDIFYTK-----T 221
DB 1 VYHC--APGGMVPMWHTSGMNGAIWVLPBDGLKDEKQGPLTYDKIYVGEOPFYPKDEA 58
QY 223 ---GKYGEQGLQPFDMKEAIRD---DAEYVFNKSVGALGTENALKAKEVETVRLFVNGG 277
DB 59 GNYKKTETGEAYEDVAKMRITLPTHTYFNGVVGALTDHDLTAAGV--RVLVHSA 116
QY 279 NLTSFHHVIGEIFDKVHFEKGKGNHN-----IQTLIPAGGAATTEFKVDVPGDYVLVD 334
DB 117 NADTRPHLIGHGDYV---GATGKFPNPPDLQDETWLIPGTAGAFYTRQPGVAYAVN 173
QY 334 AIFRA 339
DB 176 NMLEA 180

RESULT 23
ID 093TN4 PRELIMINARY; PRT; 180 AA.
AC 093TN4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative dissimilatory nitrite reductase (Fragment).
OS Nitrosomonas sp. URW.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_Taxid=159352;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=URW;
RC MEDLINE=21218534; PubMed=11319103;
RA Casciotti K.L., Ward B.B.;
RT "Dissimilatory nitrite reductase genes from autotrophic ammonia-
oxidizing bacteria."
RL Appl. Environ. Microbiol. 67:2213-2221 (2001).
DR EMBL; AF339046; AAK53360.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1
FT 180
SQ SEQUENCE 180 AA; 19540 MW; BB7CA048F7CD6D29 CRC64;

Query Match
Best Local Similarity 35.1%; Pred. No. 7.1e-08;

```


Matches 65; Conservative 24; Mismatches 60; Indels 27; Gaps 8;

QY 185 VGHINANGVGLIVKEKGLP-----KVDKEYVYVGGDFTT-----KCKY-----GE 227
 DB 3 VPMHVSGMGAVVLPREGLKQAGKALYDRAVYIGEHLVPRDEKGYRYSFG 62
 QY 228 OGQAPPD-MEKAREDEYVFNVSVALTGENALKAKVETVRLFEVNGSPNLTSSFHV 286
 DB 63 AYVDYTMELMKLI---PTHVFNKQALGKAMTKVETV--LVHSQNRDSRPHL 117
 QY 287 IGEIFDKVHEGEGKGENH---NIQTLLIPAGAAITEFKVDVPGDYLVDHAI 339
 DB 118 IGHGHVH-VETGKFARPREVDLETWFIKSGSAGALYTRPGIYAVVHNLIEA 172

RESULT 27

Q8RLT1 PRELIMINARY; PRT; 179 AA.

ID Q8RLT1
 AC Q8RLT1;
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-OCT-2002 (T-EMBLrel. 21, Last sequence update)
 DE Putative dissimilatory nitrite reductase (Fragment).
 GN NTRK.
 OS Ochrobactrum sp. 3CB4.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Ochrobactrum.
 OC NCBI_TaxID=126140;
 RN 1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=3CB4;
 RC Song B., Ward B.B.;
 RA "Nitrite reductase genes in halobenzoate degrading denitrifying
 RT bacteria.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY078250; AAL82509.1;
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.
 FT NON_TER 1 1
 FT NON_TER 179 179
 SQ SEQUENCE 179 AA; 19587 MW; 021DDB57B2A88A4F CRC64;

Query Match 8.5%; Score 223; DB 2; Length 179;
 Best Local Similarity 36.6%; Pred. No. 1.7e-07;
 Matches 67; Conservative 23; Mismatches 69; Indels 24; Gaps 8;

QY 177 VYHCAVAVGM--HINANGVGLIVKEKGLP-----KVDKEYVYVGGDFTT-----K 222
 DB 1 VYHC--APGKVPWVHVSGMGALIMVLPREGLHDKGKALTYDKIYVVGEDDFVPRDEN 58
 QY 223 GKQGE---OGIOPDMKAIKRE-DAEVVFNVSVALTGENALKAKVETVRLFEVNGSP 278
 DB 59 GKAKKYPAGDPAVEDYVAKMTLTPTVAVENGAGALGDKAMPAAGE--KVLIVSQA 116
 QY 279 NLTSFVIGIEIFDKVHEGEGKGENHNI--QTLLIPAGAAITEFKVDVPGDYLVDHAI 336
 DB 117 NRDRPHLIGHGHVYVATGKFNTPPVDDETWFIKSGSAGALYTRPGIYAVVHNL 176

QY 337 FRA 339
 DB 177 IEA 179

RESULT 28

Q9K098 PRELIMINARY; PRT; 138 AA.

ID Q9K098
 AC Q9K098;
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Cytochrome, putative.
 GN NMB0717.
 OS Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OC NCBI_TaxID=491.
 RN 1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scariato V., Masignani V., Pizsa M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58.";
 RL Science 287:1809-1815 (2000).
 DR EMBL; AE002426; AAF41130.1; -.
 DR HSSP; P04164; 1C52.
 DR TIGR; NMB0717; -.
 DR InterPro; IPR000345; Cytochrome_heme_bind.
 DR InterPro; IPR003088; Cytochrome_heme_bind.
 DR Pfam; PF00034; Cytochrome_c; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KM Complete proteome.
 SQ SEQUENCE 138 AA; 14715 MW; 947903CE2CBE6F23 CRC64;

Query Match 8.5%; Score 222; DB 16; Length 138;
 Best Local Similarity 42.7%; Pred. No. 1.3e-07;
 Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

QY 398 GKATYDNCACQHPDGKGVNPAFPLANSDYIANDHARAASIVANGSGKITVNGOYE 457
 DB 35 GKATYDNCACQHPDGKGVNPAFPLANSDYIANDHARAASIVANGSGKITVNGOYE 94
 QY 458 SWPAPALSDQOIANYITVTLNLSFGKSGQSLADYAKAKKTK 500
 DB 95 GFWPAPALSDQOIANYITVTLNLSFGKSGQSLADYAKAKKTK 137

RESULT 29

Q9JVA7 PRELIMINARY; PRT; 163 AA.

ID Q9JVA7
 AC Q9JVA7;
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DE Putative cytochrome C.
 GN NMA0925.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OC NCBI_TaxID=65699;
 RN 1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=22491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
 RA Jagers K., Leach S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.,
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491.";
 RL Nature 404:502-506 (2000).
 DR EMBL; AL162754; CAB84197.1; -.
 DR HSSP; P04164; 1C52.
 DR InterPro; IPR000345; Cytochrome_heme_bind.
 DR InterPro; IPR003088; Cytochrome_heme_bind.
 DR Pfam; PF00034; Cytochrome_c; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.

Db 1 VVHC--APPGMVMHVSGNGAIWLPREGLDGKALTYDKIYVGEODFYVPRDD 58
 QY 223 ---GKYGEGLQFDEMEKAI--DAEYVFNVSVALTGEMALKAVGETVRLFVNGGP 278
 Db 59 GKFKYDAPDADVEDVYKVRITLPTTHVFNVSVALTGEMALKAVGETVRLFVNGGP 111
 QY 279 NLTSSFFHVGELFDKVFEGGKGE-----NHNTQTLIPAGGAITEFVNDVP 326
 Db 112 ----VHSQANRTRPHLKGCHDDYWSGTCKFTPTPDVDETFFIGGAAGAFYFGQP 166
 QY 327 GDYVLVDHAFRA 339
 Db 167 GIYAVYVNHNLIEA 179

RESULT 33
 Q8RT15 PRELIMINARY; PRT; 180 AA.
 ID Q8RT15; 180 AA.
 AC Q8RT15; 180 AA.
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Copper-containing nitrite reductase (Fragment).
 GN N1RK.
 OS Azospirillum dobereineriae.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Rhodospirillaceae; Azospirillum.
 OX NCBI_TaxID=92933;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN=GSF 71T;
 RA Roesch C., Mergel A., Bothe H.;
 RT "biodiversity of denitrifying and dinitrogen fixing bacteria in an acid forest soil";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY072263; AL73092.1;
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.
 FT NON_TER 1 1
 FT NON_TER 180 180
 SQ SEQUENCE 180 AA; 19881 MW; 9A451553E7D6EFC CRC64;

Query Match 8.3%; Score 219; DB 2; Length 180;
 Best Local Similarity 33.7%; Pred. No. 3.1e-07;
 Matches 66; Conservative 26; Mismatches 56; Indels 48; Gaps 10;

QY 176 VYHCAPVGM--HIANGYGLIYVPEKGLP-----KVDKEYYVNGDFY---T 221
 Db 1 FVYHC--APPGMIPMHVSGNGLRIHWLPREGLDGKALTYDKIYVGEODFYVPRDD 58
 QY 222 KGYGE-----OGIOPDEMEKAI--DAEYVFNVSVALTGEMALKAVGETVRLFVNGGP 266
 Db 59 NGKXKDESAAGDSYADTVENKGLVP-----THVFNVSVALTGEMALKAVGETVRLFVNGGP 107
 QY 267 ETVRLFVNGGPNLTSSFFHVGELFDKVFEGGKGE---HNTQTLIPAGGAITEFVNDVP 323
 Db 108 ETV-LFI-HSANDTRPHLKGCHDDYWSGTCKFTPTPDVDETFFIGGAAGAFYFGQP 164
 QY 324 DVEGDYVLVDHAFRA 339
 Db 165 HOPGIYAVYVNHNLIEA 180

RESULT 34
 Q8KPL0 PRELIMINARY; PRT; 172 AA.
 ID Q8KPL0; 172 AA.
 AC Q8KPL0; 172 AA.
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Nitrite reductase (Fragment).
 GN N1RK.
 OS uncultured bacterium.

OC Bacteria; environmental samples.
 OX NCBI_TaxID=77133;
 RN 11)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21914100; PubMed=11916709;
 RA Prieme A., Braker G., Tiedje J.M.;
 RT "Diversity of nitrite reductase (nirK and nirS) gene fragments in forested upland and wetland soils";
 RL Appl. Environ. Microbiol. 68:1893-1900 (2002).
 DR EMBL; AY121544; AA94075.1;
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.
 FT NON_TER 1 1
 FT NON_TER 172 172
 SQ SEQUENCE 172 AA; 19273 MW; E95D4BB6340A544 CRC64;

Query Match 8.2%; Score 216.5; DB 2; Length 172;
 Best Local Similarity 35.2%; Pred. No. 4.3e-07;
 Matches 62; Conservative 27; Mismatches 60; Indels 27; Gaps 8;

QY 185 VGMHTANGYGLIYVPEKGLP-----KVDKEYYVNGDFY---KGY-----GE 227
 Db 3 VMHVHVSCHGTVVWLPREGLDGKALTYDKIYVGEODFYVPRDD 58
 QY 228 OGLOPDEMEKAI--DAEYVFNVSVALTGEMALKAVGETVRLFVNGGPNLTSSFFH 286
 Db 63 SFADTTEVVRKLI-----PTHVFNVSVALTGEMALKAVGETVRLFVNGGPNLTSSFFH 117
 QY 287 IGEIPDKVFEGGKGE---HNTQTLIPAGGAITEFVNDVPDGYVLVDHAFRA 339
 Db 118 IGHGHDYV-WETGKFNPNPQKLETFIRGSAAGALYTEREPGIYAVYVNHNLIEA 172

RESULT 35
 Q8RLT0 PRELIMINARY; PRT; 179 AA.
 ID Q8RLT0; 179 AA.
 AC Q8RLT0; 179 AA.
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Putative dissimilatory nitrite reductase (Fragment).
 GN N1RK.
 OS Ochrobactrum sp. 3CB5.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Ochrobactrum.
 OX NCBI_TaxID=126141;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN=3CB5;
 RA Song B., Ward B.B.;
 RT "Nitrite reductase genes in halobenzate degrading denitrifying bacteria";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY078251; AA82510.1;
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.
 FT NON_TER 1 1
 FT NON_TER 179 179
 SQ SEQUENCE 179 AA; 19512 MW; C51DC3A43FEE34A8 CRC64;

Query Match 8.2%; Score 216; DB 2; Length 179;
 Best Local Similarity 36.1%; Pred. No. 4.9e-07;
 Matches 66; Conservative 23; Mismatches 70; Indels 24; Gaps 8;

QY 177 VYHCAPVGM--HIANGYGLIYVPEKGLP-----KVDKEYYVNGDFY---K 222
 Db 1 VVHC--APPGMVMHVSGNGAIWLPREGLDGKALTYDKIYVGEODFYVPRDD 58
 QY 223 GKXGE---OGLOFDEMEKAI--DAEYVFNVSVALTGEMALKAVGETVRLFVNGGP 278
 Db 59 GKFKYDAPDADVEDVYKVRITLPTTHVFNVSVALTGEMALKAVGETVRLFVNGGP 116
 QY 279 NLTSSFFHVGELFDKVFEGGKGE---HNTQTLIPAGGAITEFVNDVPDGYVLVDHAFRA 336

Db 117 NR0TRPHLIGHGHDYVWATGKENTPPDVDOETWIFGGAAGAAFTPOGIIYAVVNHNL 176
 QY 337 FRA 339
 Db 177 IEA 179

RESULT 36

Q8RLT2 PRELIMINARY; PRT; 179 AA.

AC Q8RLT2;
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-OCT-2002 (T-EMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Putative dissimilatory nitrite reductase (Fragment).
 GN NIRK.

OS Ochrobactrum sp. 2FBI0.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Ochrobactrum.

OX NCBI_TaxID=126135;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2FBI0.

RA Song B., Ward B.B.;
 RT "Nitrite reductase genes in halobenzoeate degrading denitrifying
 bacteria."

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY078249; AAL82508.1; -

DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.

FT NON_TER 1 1
 FT NON_TER 179 179

SQ SEQUENCE 179 AA; 19606 MW; 021DDB57B2BDA4F CRC64;

Query Match 8.2%; Score 215; DB 2; Length 179;
 Best Local Similarity 36.1%; Pred. No. 5.8e-07;
 Matches 66; Conservative 23; Mismatches 70; Indels 24; Gaps 8;

QY 177 VYHCAVAPVGM--HIANGVYGLIVEPKKGLP-----KVDKEYVYVMOGDFTY----K 222
 Db 1 VYHC-APPGWPMVHVSGMNGAIMVLPREGLDHGKALTYKIIYVGEODTYVRDEN 58
 QY 223 GRYGE---OGLQPFDEKAIIR-DAEVYVNSVGLATGENALKAKVGETVRLFVNGNGP 278
 Db 59 GKXKYEAFGDAVEDTVKVRRTLPYHVNGVGLTGPKMTAVGE--KYLIVHQA 116
 QY 279 NLTSSRHVIGEIFDKVHFEGGKGENNT--QTLIPAGGAITTEFKVDVPGDYVLVDHAI 336
 Db 117 NR0TRPHLIGHGHDYVWATGKENTPPDVDOETWIFGGAAGAAFTPOGIIYAVVNHNL 176
 QY 337 FRA 339
 Db 177 IEA 179

RESULT 37

Q8KPM6 PRELIMINARY; PRT; 172 AA.

AC Q8KPM6;
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Nitrite reductase (Fragment).
 GN NIRK.

OS uncultured bacterium.
 OC Bacteria; environmental samples.
 OC NCBI_TaxID=77133;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21914100; PubMed=11916709;

RA Prieme A., Braker G., Tiedje J.M.;
 RT "Diversity of nitrite reductase (nirK and nirs) gene fragments in

RT forested upland and wetland soils."
 RL Appl. Environ. Microbiol. 68:1893-1900 (2002).
 DR EMBL; AY121527; AAM94058.1; -
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.
 FT NON_TER 1 1
 FT NON_TER 172 172

SQ SEQUENCE 172 AA; 19069 MW; B477803B4AFCE50F CRC64;

Query Match 8.1%; Score 213.5; DB 2; Length 172;
 Best Local Similarity 35.1%; Pred. No. 6.9e-07;
 Matches 61; Conservative 27; Mismatches 63; Indels 23; Gaps 7;

QY 185 VGMHANGVYGLIVEPKKGLP-----KVDKEYVYVMOGDFTY-----TKGKYGEQGL 230
 Db 3 VPMHVSGMGTVMVLPRLDGLNKGESVRYDRIYVGENEFYIPDETGAAYKVEISLG- 61
 QY 231 QEPDEKAIRED--AEVYVNSVGLATGENALKAKVGETVRLFVNGNGPNTSSRHVIG 288
 Db 62 ESYDETMAVMRGLIPHVHVNGKVGSLTGENALKSKVGETV--LIVHQAANDTRXHLIG 119
 QY 289 EIFDKVHFEGGKGEN--HNIQTLIPAGGAITTEFKVDVPGDYVLVDHAI FRA 339
 Db 120 GHGDIY-WEQGFANPPAKDLETWIFRGSGAGAGLYTFLOPGIIYAVVNHNL IEA 172

RESULT 38

Q8KPM5 PRELIMINARY; PRT; 172 AA.

AC Q8KPM5;
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Nitrite reductase (Fragment).

OS uncultured bacterium.
 OC Bacteria; environmental samples.
 OC NCBI_TaxID=77133;

RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=21914100; PubMed=11916709;
 RA Prieme A., Braker G., Tiedje J.M.;

RT "Diversity of nitrite reductase (nirK and nirs) gene fragments in
 forested upland and wetland soils."

RL Appl. Environ. Microbiol. 68:1893-1900 (2002).
 DR EMBL; AY121528; AAM94059.1; -

DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.

FT NON_TER 1 1
 FT NON_TER 172 172

SQ SEQUENCE 172 AA; 19275 MW; 9DCC6990239F1C0A CRC64;

Query Match 8.1%; Score 211.5; DB 2; Length 172;
 Best Local Similarity 34.7%; Pred. No. 9.4e-07;
 Matches 61; Conservative 27; Mismatches 61; Indels 27; Gaps 8;

QY 185 VGMHANGVYGLIVEPKKGLP-----KVDKEYVYVMOGDFTY-----KGK-----GE 227
 Db 3 VPMHVSGMGTVMVLPRLDGLNKGESVRYDRIYVGENEFYIPDETGAAYKVEISLG 62
 QY 228 OGLQPFDEKAIRED--AEVYVNSVGLATGENALKAKVGETVRLFVNGNGPNTSSRHVIG 286
 Db 63 SYADTTEVMRKLI---PQHVVNGVGLTGNAMTSKVGELVMLV--HSQANRDTXRPHL 117
 QY 287 ICEIFDKVHFEGGKGEN--HNIQTLIPAGGAITTEFKVDVPGDYVLVDHAI FRA 339
 Db 118 IGHGDIY-WEAGKKNPPOKDLTETWIFVFGSGAGALYTFRBPFIYAVVNHNL IEA 172

RESULT 39

Q8KPK9 PRELIMINARY; PRT; 172 AA.

AC Q8KPK9;
 QY Q8KPK9;

DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Nitrite reductase (Fragment).
 GN NITRK.
 OS uncultured bacterium.
 OC Bacteria; environmental samples.
 OX NCBI_TaxID=77133;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21914100; Pubmed=11916709;
 RA Prieme A., Braker G., Tiedje J.M.;
 RT "Diversity of nitrite reductase (nitr and nirs) gene fragments in
 RT forested upland and wetland soils."
 RL Appl. Environ. Microbiol. 68:1893-1900(2002).
 DR EMBL; AY121545; AAM94076.1; -
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.
 DR NON_TER 1
 FT 172 172
 SQ SEQUENCE 172 AA; 19137 MW; 06E260DBA3C5B345 CRC64;

Query Match 8.1%; Score 211.5; DB 2; Length 172;
 Best Local Similarity 34.7%; Pred. No. 9.4e-07;
 Matches 61; Conservative 27; Mismatches 61; Indels 27; Gaps 8;

QY 185 VGMHANGMYGLIVERKGLP-----KVDKEYVMGDFYT---KGKY-----GE 227
 DB 3 VPMHVVSGMGTVMVPRDGLKDAEGKALRYDKIFYIGENDLVYPKDNKGKTYETIGE 62
 QY 228 OGLOPFD-MEKAIRDAEYVFNQSVGALTGEMALAKAVGETVRLFVNGGPNLTSSFHV 286
 DB 63 SYADTTEVVRKLI---PTHVFNGRVGSLTGKAMTSKVGETVMLV--HSQANRDTDRPHL 117
 QY 287 TGEIFDKVHFEKGKGEN--HNIOITLIPAGGAATTEFKVDVPGDVLVDHAIFFA 339
 DB 118 IGHGDHV-WEAGKFNPNPQKDLFTWFRGSGAGALYTFREPGIYAYVNHNLIEA 172

RESULT 40

Q8KPM9 PRELIMINARY; PRT; 172 AA.
 AC Q8KPM9;
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Nitrite reductase (Fragment).
 GN NITRK.
 OS uncultured bacterium.
 OC Bacteria; environmental samples.
 OX NCBI_TaxID=77133;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21914100; Pubmed=11916709;
 RA Prieme A., Braker G., Tiedje J.M.;
 RT "Diversity of nitrite reductase (nitr and nirs) gene fragments in
 RT forested upland and wetland soils."
 RL Appl. Environ. Microbiol. 68:1893-1900(2002).
 DR EMBL; AY121521; AAM94052.1; -
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 1.
 DR NON_TER 1
 FT 172 172
 SQ SEQUENCE 172 AA; 19146 MW; 1B490D6A18695C45 CRC64;

Query Match 8.1%; Score 211.5; DB 2; Length 172;
 Best Local Similarity 34.1%; Pred. No. 9.4e-07;
 Matches 60; Conservative 28; Mismatches 61; Indels 27; Gaps 8;

QY 185 VGMHANGMYGLIVERKGLP-----KVDKEYVMGDFYT---KGKY-----GE 227
 DB 3 VPMHVVSGMGTVMVPRDGLKDAEGKALRYDKIFYIGENDLVYPKDNKGKTYETIGE 62

QY 228 OGLOPFD-MEKAIRDAEYVFNQSVGALTGEMALAKAVGETVRLFVNGGPNLTSSFHV 286
 DB 63 SYADTTEVVRKLI---PTHVFNGRVGSLTGKAMTSKVGETVMLV--HSQANRDTDRPHL 117
 QY 287 TGEIFDKVHFEKGKGEN--HNIOITLIPAGGAATTEFKVDVPGDVLVDHAIFFA 339
 DB 118 IGHGDHV-WEAGKFNPNPQKDLFTWFRGSGAGAAXTFREPGIYAYVNHNLIEA 172

Search completed: August 27, 2003, 18:31:55
 Job time : 106 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:33:47 ; Search time 104 Seconds
(without alignments)
1245.600 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502
Sequence: 1 MSKPTLIKTTLLCAASALML.....NKGQLSADDAVAKAKTKPEN 502

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp Unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	6.8	186	16	Q9JTB8
2	34	6.8	390	16	Q9JTE1
3	19	3.8	510	16	Q8XPY3
4	16	3.2	159	2	Q8GBL8
5	12	2.4	94	14	Q8BFN8
6	12	2.4	95	14	Q8BFP2
7	12	2.4	99	14	Q8BFL7
8	12	2.4	102	14	Q8BFN9
9	12	2.4	102	14	Q8BFN5
10	12	2.4	102	14	Q8BFN3
11	12	2.4	102	14	Q8BFN0
12	12	2.4	102	14	Q8BFN8
13	12	2.4	102	14	Q8BFM2
14	12	2.4	102	14	Q8BFL4
15	12	2.4	102	14	Q8BFX9
16	12	2.4	102	14	Q8BFP6

17	12	2.4	103	14	Q8BFP0	Q8BFP0 uncultured
18	12	2.4	172	2	Q87751	Q87751 hypomicrob
19	12	2.4	142	2	Q8KPL1	Q8KPL1 uncultured
20	12	2.4	360	2	Q9AQ07	Q9AQ07 alcaligenes
21	12	2.4	486	2	Q8KKH4	Q8KKH4 hypomicrob
22	11	2.2	102	14	Q8BFL0	Q8BFL0 uncultured
23	11	2.2	102	14	Q8BFL6	Q8BFL6 uncultured
24	10	2.0	361	1	Q9PPH9	Q9PPH9 haloarcula
25	9	1.8	172	2	Q8KPL6	Q8KPL6 uncultured
26	9	1.8	428	16	Q8CX87	Q8CX87 oceanobacil
27	8	1.6	174	16	Q8X8W0	Q8X8W0 escherichia
28	8	1.6	131	2	Q8KPL1	Q8KPL1 uncultured
29	8	1.6	172	2	Q8KPL6	Q8KPL6 uncultured
30	8	1.6	252	17	Q8TSG4	Q8TSG4 methanosarc
31	8	1.6	252	17	Q8PVL5	Q8PVL5 methanosarc
32	8	1.6	255	10	Q9FFM7	Q9FFM7 arabidopsis
33	8	1.6	321	16	Q8EUC9	Q8EUC9 mycoplasma
34	8	1.6	360	2	Q86601	Q86601 alcaligenes
35	8	1.6	483	16	Q9KLC4	Q9KLC4 neisseria m
36	8	1.6	494	16	Q9JX77	Q9JX77 neisseria m
37	8	1.6	509	16	Q98112	Q98112 rhizobium 1
38	8	1.6	514	10	Q9AV50	Q9AV50 oryza sativ
39	8	1.6	634	10	Q9LZB8	Q9LZB8 arabidopsis
40	8	1.6	915	5	Q93538	Q93538 caenorhabdi
41	8	1.6	995	16	Q866F8	Q866F8 bifidobacte
42	7	1.4	61	16	Q8KE17	Q8KE17 chlorobium
43	7	1.4	77	9	Q9MCT9	Q9MCT9 bacterioph
44	7	1.4	90	5	Q97102	Q97102 drosophila
45	7	1.4	98	16	Q8U8F9	Q8U8F9 agrobacteri
46	7	1.4	101	10	Q8GVQ4	Q8GVQ4 oryza sativ
47	7	1.4	102	14	Q8BFL1	Q8BFL1 uncultured
48	7	1.4	102	14	Q8BFN6	Q8BFN6 uncultured
49	7	1.4	102	14	Q8BFM9	Q8BFM9 uncultured
50	7	1.4	102	14	Q8BFM5	Q8BFM5 uncultured
51	7	1.4	102	14	Q8BFL4	Q8BFL4 uncultured
52	7	1.4	102	14	Q8BFL5	Q8BFL5 uncultured
53	7	1.4	102	14	Q8BFL2	Q8BFL2 uncultured
54	7	1.4	103	14	Q8BFLK	Q8BFLK uncultured
55	7	1.4	103	11	Q8VDZ2	Q8VDZ2 mus musculu
56	7	1.4	104	2	Q53039	Q53039 rhodococcus
57	7	1.4	104	15	Q76163	Q76163 human timm
58	7	1.4	112	2	Q9L477	Q9L477 uncultured
59	7	1.4	112	2	Q9L484	Q9L484 uncultured
60	7	1.4	112	2	Q9L483	Q9L483 uncultured
61	7	1.4	112	2	Q9L478	Q9L478 uncultured
62	7	1.4	112	2	Q9L480	Q9L480 uncultured
63	7	1.4	118	11	Q92364	Q92364 rattus norv
64	7	1.4	133	10	Q41939	Q41939 arabidopsis
65	7	1.4	135	10	Q43037	Q43037 populus del
66	7	1.4	139	16	Q8FOT3	Q8FOT3 leptospira
67	7	1.4	142	2	Q87684	Q87684 alcaligenes
68	7	1.4	142	2	Q87686	Q87686 achromobact
69	7	1.4	142	2	Q9R801	Q9R801 achromobact
70	7	1.4	157	2	Q9X4L1	Q9X4L1 paracoccus
71	7	1.4	159	2	Q8GBL3	Q8GBL3 uncultured
72	7	1.4	160	10	Q39454	Q39454 coriulus ave
73	7	1.4	161	16	Q92HF3	Q92HF3 rickettsia
74	7	1.4	165	16	Q8CU81	Q8CU81 staphylococ
75	7	1.4	165	16	Q8CU70	Q8CU70 staphylococ
76	7	1.4	167	5	Q9V218	Q9V218 drosophila
77	7	1.4	168	2	Q87591	Q87591 streptomyces
78	7	1.4	169	2	Q8KPL0	Q8KPL0 uncultured
79	7	1.4	169	11	Q8C329	Q8C329 mus musculu
80	7	1.4	170	5	Q817G2	Q817G2 caenorhabdi
81	7	1.4	170	16	Q8XUF5	Q8XUF5 raietonia s
82	7	1.4	172	2	Q9R363	Q9R363 pseudomonas
83	7	1.4	172	2	Q9PW18	Q9PW18 pseudomonas
84	7	1.4	172	2	Q8KPK0	Q8KPK0 uncultured
85	7	1.4	172	2	Q9R433	Q9R433 pseudomonas
86	7	1.4	172	2	Q8KPL1	Q8KPL1 uncultured
87	7	1.4	172	2	Q8KPL0	Q8KPL0 uncultured
88	7	1.4	172	2	Q9RBY3	Q9RBY3 pseudomonas
89	7	1.4	172	2	Q8KPM7	Q8KPM7 uncultured

90	7	1.4	172	2	Q9R434	Q9r434 pseudomonas	163	7	1.4	268	2	069346	069346 pseudomonas
91	7	1.4	172	2	Q9R450	Q9r450 pseudomonas	164	7	1.4	268	2	Q9ERT3	Q9ert3 pseudomonas
92	7	1.4	172	2	Q8Kpj7	Q8Kpj7 uncultured	165	7	1.4	268	2	Q9FAF9	Q9faf9 pseudomonas
93	7	1.4	172	2	Q9R449	Q9r449 pseudomonas	166	7	1.4	269	2	Q9FAL1	Q9fal1 pseudomonas
94	7	1.4	172	2	Q9RBV2	Q9rbv2 pseudomonas	167	7	1.4	269	2	Q9FAL6	Q9fal6 pseudomonas
95	7	1.4	172	2	Q9R9F5	Q9r9f5 pseudomonas	168	7	1.4	269	2	Q9FAL2	Q9fal2 pseudomonas
96	7	1.4	172	2	Q9R2S2	Q9r2s2 pseudomonas	169	7	1.4	269	2	Q9ERT6	Q9ert6 pseudomonas
97	7	1.4	172	2	Q9R9F8	Q9r9f8 pseudomonas	170	7	1.4	269	2	Q9FAJ1	Q9faj1 pseudomonas
98	7	1.4	172	2	Q8KPK3	Q8Kpk3 uncultured	171	7	1.4	269	2	Q9ERT7	Q9ert7 pseudomonas
99	7	1.4	173	2	Q8RT14	Q8rt14 azospirillum	172	7	1.4	269	2	Q9ERT1	Q9ert1 pseudomonas
100	7	1.4	173	16	Q8XY56	Q8xy56 ralstonia s	173	7	1.4	269	2	Q9FAH8	Q9fah8 pseudomonas
101	7	1.4	174	2	Q9WJ1	Q9wj1 pseudomonas	174	7	1.4	269	2	Q9FAJ5	Q9faj5 pseudomonas
102	7	1.4	175	16	Q92IT3	Q92it3 rickettsia	175	7	1.4	269	2	Q9FAK7	Q9fak7 pseudomonas
103	7	1.4	179	2	Q8RLS6	Q8rls6 pseudomonas	176	7	1.4	269	2	Q9FAH9	Q9fah9 pseudomonas
104	7	1.4	179	2	Q93TN1	Q93tn1 nitrosomona	177	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
105	7	1.4	180	2	Q8RT15	Q8rt15 azospirillum	178	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
106	7	1.4	180	2	Q93TN6	Q93tn6 nitrosomona	179	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
107	7	1.4	180	2	Q93TN3	Q93tn3 nitrosomona	180	7	1.4	269	2	Q9ERT7	Q9ert7 pseudomonas
108	7	1.4	180	2	Q93TN2	Q93tn2 nitrosomona	181	7	1.4	269	2	Q9FAJ6	Q9faj6 pseudomonas
109	7	1.4	180	2	Q93TN4	Q93tn4 nitrosomona	182	7	1.4	269	2	Q9FAJ6	Q9faj6 pseudomonas
110	7	1.4	183	5	Q9VW10	Q9vw10 nitrosomona	183	7	1.4	269	2	Q9ERT9	Q9ert9 pseudomonas
111	7	1.4	185	5	Q9VIZ1	Q9vizi dtrosophila	184	7	1.4	269	2	Q69344	Q69344 pseudomonas
112	7	1.4	187	16	Q8KEZ8	Q8kez8 chlorobium	185	7	1.4	269	2	Q9ERT2	Q9ert2 pseudomonas
113	7	1.4	188	10	Q9FVY0	Q9fvy0 oryza sativ	186	7	1.4	269	2	Q9FAI2	Q9fai2 pseudomonas
114	7	1.4	188	10	Q8H8Y9	Q8h8y9 oryza sativ	187	7	1.4	269	2	Q9ERT8	Q9ert8 pseudomonas
115	7	1.4	191	16	Q8ZFS5	Q8zfs5 yersinia pe	188	7	1.4	269	2	Q9FAG1	Q9fag1 pseudomonas
116	7	1.4	192	5	Q23065	Q23065 caenorhabdi	189	7	1.4	269	2	Q9ERT5	Q9ert5 pseudomonas
117	7	1.4	193	16	Q8XET8	Q8xet8 salmonella	190	7	1.4	269	2	Q9FAL0	Q9fal0 pseudomonas
118	7	1.4	193	16	Q8XDR5	Q8xd5 escherichia	191	7	1.4	269	2	Q69343	Q69343 pseudomonas
119	7	1.4	195	2	Q54673	Q54673 lactococcus	192	7	1.4	269	2	Q9FAJ2	Q9faj2 pseudomonas
120	7	1.4	196	2	Q8GC69	Q8gc69 escherichia	193	7	1.4	269	2	Q9FAK9	Q9fak9 pseudomonas
121	7	1.4	196	2	Q8GC68	Q8gc68 escherichia	194	7	1.4	269	2	Q9FAJ0	Q9faj0 pseudomonas
122	7	1.4	200	2	Q8VNR7	Q8vnr7 glucinobact	195	7	1.4	269	2	Q9EUB0	Q9eub0 pseudomonas
123	7	1.4	201	2	Q50132	Q50132 mycobacteri	196	7	1.4	269	2	Q9FAI4	Q9fai4 pseudomonas
124	7	1.4	202	2	Q54675	Q54675 lactococcus	197	7	1.4	270	2	Q69339	Q69339 pseudomonas
125	7	1.4	202	16	Q9S2G6	Q9s2g6 streptococ	198	7	1.4	270	2	Q9FAK6	Q9fak6 pseudomonas
126	7	1.4	208	2	Q9AJ86	Q9aj86 lactococcus	199	7	1.4	271	2	Q9FAK8	Q9fak8 pseudomonas
127	7	1.4	209	2	Q9AC98	Q9ac98 lactococcus	200	7	1.4	276	16	Q8YVX1	Q8yvx1 anabaena sp
128	7	1.4	209	2	Q93K29	Q93k29 lactococcus	201	7	1.4	278	16	Q8NQD7	Q8nd7 corynebacte
129	7	1.4	210	2	Q9JN26	Q9jnz6 lactococcus	202	7	1.4	284	16	Q8G264	Q8g264 brucella su
130	7	1.4	210	2	Q9REY1	Q9rey1 lactococcus	203	7	1.4	285	2	Q52264	Q52264 pseudomonas
131	7	1.4	211	10	Q9SST3	Q9sse3 arabidopsis	204	7	1.4	289	10	Q8S5B6	Q8sb6 oryza sativ
132	7	1.4	211	5	Q9S5J3	Q9sj53 arabidopsis	205	7	1.4	293	5	P91356	P91356 caenorhabdi
133	7	1.4	212	5	P91355	P91355 caenorhabdi	206	7	1.4	293	16	Q9CIJ3	Q9cij3 lactococcus
134	7	1.4	213	2	Q8VMA1	Q8vma1 streptococ	207	7	1.4	296	16	Q87127	Q87127 pseudomonas
135	7	1.4	214	2	Q8L0W1	Q8l0w1 lactococcus	208	7	1.4	298	4	Q9NZE1	Q9nze1 homo sapien
136	7	1.4	215	10	Q43719	Q43719 lycopersico	209	7	1.4	299	4	Q8N4S8	Q8nas8 homo sapien
137	7	1.4	216	5	Q20601	Q20601 caenorhabdi	210	7	1.4	299	10	Q9SLV2	Q9slv2 arabidopsis
138	7	1.4	218	10	Q8S1D3	Q8s1d3 oryza sativ	211	7	1.4	299	16	Q8U8U3	Q8u8u3 agrobacteri
139	7	1.4	218	16	Q8XB27	Q8xb27 escherichia	212	7	1.4	299	16	Q8U9Y6	Q8u9y6 agrobacteri
140	7	1.4	222	1	Q9UMN5	Q9umn5 methanother	213	7	1.4	300	5	Q9V7H9	Q9v7h9 dtrosophila
141	7	1.4	228	5	Q9VB99	Q9vb99 dtrosophila	214	7	1.4	300	16	Q9RK57	Q9rk57 streptococ
142	7	1.4	228	10	Q9FVL8	Q9fvl8 nicotiana a	215	7	1.4	301	5	Q8IKH7	Q8ikh7 habesla div
143	7	1.4	230	15	Q9IE91	Q9ie91 human immun	216	7	1.4	305	16	Q8UD20	Q8ud20 agrobacteri
144	7	1.4	232	16	Q8KAW5	Q8kaws chlorobium	217	7	1.4	309	16	Q9BHV4	Q9bhv4 campylobact
145	7	1.4	239	12	Q9DUQ4	Q9duq4 rice yellow	218	7	1.4	309	16	Q8CR55	Q8cr55 staphylococ
146	7	1.4	245	16	Q9CDU7	Q9cdz7 lactococcus	219	7	1.4	310	5	Q8IKH4	Q8ikh4 habesla div
147	7	1.4	247	10	Q8LIG3	Q8lig3 oryza sativ	220	7	1.4	312	2	P71214	P71214 escherichia
148	7	1.4	248	5	Q8IMQ7	Q8imq7 dtrosophila	221	7	1.4	312	2	Q8S6Z5	Q8s6z5 escherichia
149	7	1.4	251	16	Q8XXV6	Q8xxv6 ralstonia s	222	7	1.4	312	16	Q8XC86	Q8xc86 escherichia
150	7	1.4	255	16	Q9PCQ2	Q9pcq2 xylella fas	223	7	1.4	313	2	Q9RZ26	Q9rz26 bordetella
151	7	1.4	257	2	Q9PF40	Q9pf40 stigmatella	224	7	1.4	313	10	Q9LQ88	Q9lqb8 arabidopsis
152	7	1.4	258	10	Q9AUD0	Q9aud0 oryza sativ	225	7	1.4	315	16	Q9PR31	Q9pr31 ureaplasma
153	7	1.4	258	17	Q9HHR7	Q9hhz7 halobacteri	226	7	1.4	318	16	Q8KD57	Q8kds7 chlorobium
154	7	1.4	261	16	Q9HY27	Q9hy27 pseudomonas	227	7	1.4	320	16	Q8IFQ4	Q8ifq4 brucella me
155	7	1.4	264	2	Q8GE45	Q8ge45 heliobacilli	228	7	1.4	321	16	Q8PFQ4	Q8pfq4 xanthomonas
156	7	1.4	267	5	Q9VSU7	Q9vsu7 dtrosophila	229	7	1.4	322	10	Q9SOL1	Q9sol1 musa acumin
157	7	1.4	268	2	Q9FAJ8	Q9faj8 pseudomonas	230	7	1.4	322	16	Q97DS2	Q97ds2 clostridium
158	7	1.4	268	2	Q9FAK2	Q9fak2 pseudomonas	231	7	1.4	323	4	Q96102	Q96102 homo sapien
159	7	1.4	268	2	Q9FAK1	Q9fak1 pseudomonas	232	7	1.4	323	16	Q9A7C4	Q9a7c4 caulobacter
160	7	1.4	268	2	Q9ERT0	Q9ert0 pseudomonas	233	7	1.4	326	10	Q9ZPK2	Q9zpk2 pieum sativ
161	7	1.4	268	2	Q9FAK3	Q9fak3 pseudomonas	234	7	1.4	326	10	Q9SD22	Q9sd22 glycine max
162	7	1.4	268	2	Q9FAK3	Q9fak3 pseudomonas	235	7	1.4	326	10	Q40316	Q40316 medicago sa

236	7	1.4	327	10	046602	046602 glycine max	309	7	1.4	512	2	0936C7	0936C7 streptomyce
237	7	1.4	328	4	043858	043858 homo sapien	310	7	1.4	512	4	08N5C0	08N5C0 homo sapien
238	7	1.4	331	4	044031	044031 ralestonia m	311	7	1.4	513	11	091V97	091V97 mus musculus
239	7	1.4	331	5	081HK2	081HK2 babesia div	312	7	1.4	513	16	08E805	08E805 shewanella
240	7	1.4	334	17	028427	028427 archaeoglob	313	7	1.4	516	16	0970P4	0970P4 streptococ
241	7	1.4	338	16	097H52	097H52 clostridium	314	7	1.4	519	10	08W0H5	08W0H5 oryza sativ
242	7	1.4	338	16	08D532	08D532 vibrio vuln	315	7	1.4	520	4	09UQ43	09UQ43 homo sapien
243	7	1.4	340	10	022317	022317 musa acum	316	7	1.4	521	11	060437	060437 cilicetidae
244	7	1.4	341	4	081W73	081W73 homo sapien	317	7	1.4	521	11	09UBR3	09UBR3 homo sapien
245	7	1.4	341	11	099L01	099L01 mus musculu	318	7	1.4	521	16	08DPR2	08DPR2 streptococ
246	7	1.4	341	16	0917A7	0917A7 pseudomonas	319	7	1.4	522	4	096CR7	096CR7 homo sapien
247	7	1.4	341	16	09AB77	09AB77 caulobacter	320	7	1.4	522	11	0923W0	0923W0 mus musculu
248	7	1.4	343	4	096A55	096A55 homo sapien	321	7	1.4	522	11	0923H3	0923H3 ractus norv
249	7	1.4	344	4	09NS73	09NS73 homo sapien	322	7	1.4	524	11	09D2N5	09D2N5 mus musculu
250	7	1.4	344	4	09BS93	09BS93 homo sapien	323	7	1.4	528	2	093R40	093R40 colwellia m
251	7	1.4	349	11	08K1A7	08K1A7 mus musculu	324	7	1.4	529	16	098CJ7	098CJ7 rhizobium l
252	7	1.4	349	11	09LTC4	09LTC4 mus musculu	325	7	1.4	535	11	088KX1	088KX1 mus musculu
253	7	1.4	352	2	052378	052378 pseudomonas	326	7	1.4	535	16	092XB1	092XB1 rhizobium m
254	7	1.4	359	5	020880	020880 caenorhabd	327	7	1.4	543	5	020720	020720 caenorhabd
255	7	1.4	363	16	08XQ33	08XQ33 ralestonia s	328	7	1.4	544	5	027659	027659 giardia lam
256	7	1.4	363	16	08KBE1	08KBE1 chlorobium	329	7	1.4	545	16	08RHNO	08RHNO fusobacteri
257	7	1.4	373	9	09AZR7	09AZR7 bacterioph	330	7	1.4	546	17	09HLE7	09HLE7 thermoplaam
258	7	1.4	373	16	09CFT0	09CFT0 laccococcus	331	7	1.4	547	2	P74913	P74913 treponema d
259	7	1.4	373	17	09V014	09V014 pyrococcus	332	7	1.4	548	4	09BWZ5	09BWZ5 homo sapien
260	7	1.4	376	16	08YBB2	08YBB2 bruceella me	333	7	1.4	552	4	09UQB8	09UQB8 homo sapien
261	7	1.4	376	16	08FX29	08FX29 bruceella su	334	7	1.4	553	4	081Z85	081Z85 homo sapien
262	7	1.4	379	2	085863	085863 sphingomona	335	7	1.4	554	4	09YJF0	09YJF0 homo sapien
263	7	1.4	380	17	0971D3	0971D3 sulfolobus	336	7	1.4	555	5	0961R9	0961R9 drosoophila
264	7	1.4	384	17	09HKY2	09HKY2 thermoplaam	337	7	1.4	558	4	08NA06	08NA06 homo sapien
265	7	1.4	388	16	08EVA2	08EVA2 mycoplasma	338	7	1.4	565	4	081W66	081W66 homo sapien
266	7	1.4	389	16	08EVA0	08EVA0 mycoplasma	339	7	1.4	567	16	09ZME2	09ZME2 helicobacte
267	7	1.4	391	3	09C1C2	09C1C2 glomerella	340	7	1.4	567	16	08XS30	08XS30 ralestonia s
268	7	1.4	391	16	09XZ23	09XZ23 thermocoga	341	7	1.4	571	17	08TXB7	08TXB7 methanopyru
269	7	1.4	393	10	09LD07	09LD07 camellia si	342	7	1.4	576	4	096W93	096W93 homo sapien
270	7	1.4	393	10	09AT55	09AT55 elaeagnus u	343	7	1.4	583	16	08RFT9	08RFT9 fusobacteri
271	7	1.4	393	10	09AT56	09AT56 elaeagnus u	344	7	1.4	586	4	09NZ29	09NZ29 homo sapien
272	7	1.4	395	10	09AT56	09AT56 elaeagnus u	345	7	1.4	587	4	09H859	09H859 ractus norv
273	7	1.4	396	10	09FPL6	09FPL6 suada mari	346	7	1.4	588	11	09D604	09D604 mus musculu
274	7	1.4	398	4	08N6H2	08N6H2 oryza sativ	347	7	1.4	588	11	08BTW8	08BTW8 mus musculu
275	7	1.4	399	4	081VQ3	081VQ3 homo sapien	348	7	1.4	594	11	08BD21	08BD21 mus musculu
276	7	1.4	399	10	08W0M3	08W0M3 oryza sativ	349	7	1.4	594	11	08BD21	08BD21 mus musculu
277	7	1.4	399	16	08ESU4	08ESU4 oceanobacil	350	7	1.4	595	10	0946Y5	0946Y5 chlamydom
278	7	1.4	403	5	09U250	09U250 caenorhabd	351	7	1.4	600	5	09V614	09V614 drosoophila
279	7	1.4	411	3	09C196	09C196 aspergillus	352	7	1.4	608	16	09HZV4	09HZV4 pseudomonas
280	7	1.4	412	12	091TM4	091TM4 tupia hepp	353	7	1.4	614	17	097B30	097B30 thermoplaam
281	7	1.4	413	5	08SS43	08SS43 entcephalito	354	7	1.4	616	2	087812	087812 pseudomonas
282	7	1.4	419	4	043611	043611 homo sapien	355	7	1.4	617	2	09EVD8	09EVD8 pseudomonas
283	7	1.4	419	10	09LUT4	09LUT4 arabidopsis	356	7	1.4	617	2	09EVB2	09EVB2 pseudomonas
284	7	1.4	420	16	0982K3	0982K3 rhizobium l	357	7	1.4	617	2	09RBY1	09RBY1 pseudomonas
285	7	1.4	420	16	09RUE8	09RUE8 deinococcus	358	7	1.4	617	2	09X3C0	09X3C0 pseudomonas
286	7	1.4	429	10	08GTL5	08GTL5 carica papa	359	7	1.4	663	16	09KR77	09KR77 vibrio chol
287	7	1.4	431	2	08GDL8	08GDL8 bacillus th	360	7	1.4	672	5	09VMU5	09VMU5 drosoophila
288	7	1.4	438	2	052582	052582 ecaphylococ	361	7	1.4	714	16	09A5S3	09A5S3 caulobacter
289	7	1.4	438	16	09KAH8	09KAH8 bacillus ha	362	7	1.4	716	16	09RZD1	09RZD1 deinococcus
290	7	1.4	438	16	099VC0	099VC0 staphylococ	363	7	1.4	722	2	09FI64	09FI64 amycolatops
291	7	1.4	438	16	08NXX8	08NXX8 staphylococ	364	7	1.4	744	13	091149	091149 nocophthalm
292	7	1.4	447	5	020703	020703 caenorhabd	365	7	1.4	747	2	09ZAX6	09ZAX6 laccococcus
293	7	1.4	450	2	059658	059658 pelobacter	366	7	1.4	747	16	09CIT3	09CIT3 laccococcus
294	7	1.4	451	2	045695	045695 burkholderi	367	7	1.4	764	2	08KFW9	08KFW9 saccharopol
295	7	1.4	455	16	08U7R7	08U7R7 agrobacteri	368	7	1.4	779	16	08YNU0	08YNU0 anabaena sp
296	7	1.4	466	3	000100	000100 aspergillus	369	7	1.4	789	17	08TXR3	08TXR3 methanopyru
297	7	1.4	472	2	08KZS7	08KZS7 acetobacter	370	7	1.4	800	12	09WAS5	09WAS5 zucchini ye
298	7	1.4	472	2	057536	057536 acetobacter	371	7	1.4	800	12	09WAS5	09WAS5 zucchini ye
299	7	1.4	473	16	08CIS9	08CIS9 streptomyce	372	7	1.4	804	2	08GCV4	08GCV4 laccobacill
300	7	1.4	479	16	08FTW0	08FTW0 corynebacte	373	7	1.4	815	13	P87376	P87376 tritriturus ca
301	7	1.4	491	5	096638	096638 trypanosoma	374	7	1.4	844	16	025856	025856 helicobacte
302	7	1.4	494	16	08D7Z7	08D7Z7 vibrio vuln	375	7	1.4	849	16	09ZJW1	09ZJW1 african swi
303	7	1.4	495	16	092UJ3	092UJ3 rhizobium m	376	7	1.4	868	12	065216	065216 african swi
304	7	1.4	501	16	09A496	09A496 caulobacter	377	7	1.4	869	5	081G55	081G55 caenorhabd
305	7	1.4	509	12	08QWJ9	08QWJ9 ectocarpus	378	7	1.4	878	16	08EM47	08EM47 mycoplasma
306	7	1.4	510	4	09BVG6	09BVG6 homo sapien	379	7	1.4	884	11	08BN06	08BN06 mus musculu
307	7	1.4	511	2	09K4U0	09K4U0 mycobacteri	380	7	1.4	884	15	08Q716	08Q716 human immun
308	7	1.4	511	5	09W2H0	09W2H0 dtrosophila	381	7	1.4	889	2	08RMW8	08RMW8 azotobacter

382	7	1.4	923	10	049003	049003	avena sativ	455	6	1.2	57	16	096MG9	096mg rhizobium 1
383	7	1.4	955	16	092625	092625	salmonella	456	6	1.2	57	16	08NPJ7	08npj7 corynebacte
384	7	1.4	970	12	091TM6	091tm6	tupia herp	457	6	1.2	58	12	08NPJ8	08npj8 heparitis c
385	7	1.4	988	11	08C035	08C035	mus musculu	458	6	1.2	58	13	013084	013084 discopge o
386	7	1.4	1012	11	08CHU1	08CHU1	mus musculu	459	6	1.2	58	16	09WYH6	09wyh6 thermocoga
387	7	1.4	1024	4	014527	014527	homo sapien	460	6	1.2	61	2	09X4G6	09x4g6 vibrio chol
388	7	1.4	1024	5	081577	081577	plasmodium	461	6	1.2	61	10	09M0F1	09m0f1 aradidopsis
389	7	1.4	1046	11	055156	055156	homo sapien	462	6	1.2	62	4	096HE3	096he3 homo sapien
390	7	1.4	1046	11	0920H8	0920H8	rattus norv	463	6	1.2	63	10	09AXV6	09axv6 brassica na
391	7	1.4	1046	11	0920H8	0920H8	mus musculu	464	6	1.2	63	16	093J37	093j37 streptomyc
392	7	1.4	1047	11	09EP81	09EP81	mus musculu	465	6	1.2	64	2	09ARF6	09arf6 shigella fl
393	7	1.4	1102	5	09LVPO	09LVPO	aradidopsis	466	6	1.2	64	10	0943H5	0943h5 oryza sativ
394	7	1.4	1132	5	0815C3	0815C3	plasmodium	467	6	1.2	64	16	08G4L2	08g4l2 bifidobacte
395	7	1.4	1164	12	09ULI1	09ULI1	homo sapien	468	6	1.2	66	10	09AXV1	09axv1 brassica ol
396	7	1.4	1164	12	089331	089331	zucchini ye	469	6	1.2	66	10	09AXV0	09axv0 brassica ol
397	7	1.4	1164	12	0886A8	0886A8	zucchini ye	470	6	1.2	67	10	09AXV2	09axv2 brassica ca
398	7	1.4	1174	5	09VXZ6	09VXZ6	dirosophila	471	6	1.2	67	16	09KRJ4	09krj4 vibrio chol
399	7	1.4	1174	11	08CHA7	08CHA7	mus musculu	472	6	1.2	68	6	09BDW4	09bdw4 sus scrofa
400	7	1.4	1219	5	09XZ09	09XZ09	dirosophila	473	6	1.2	68	13	090510	090510 figu rubrip
401	7	1.4	1232	5	08T821	08T821	dictyosteli	474	6	1.2	68	16	08KC22	08kc22 chlorobium
402	7	1.4	1248	16	08U01	08U01	agrobacteri	475	6	1.2	69	2	09L7G2	09l7g2 synchococc
403	7	1.4	1257	5	08IOH9	08IOH9	dirosophila	476	6	1.2	69	13	08AXL9	08axl9 oncorhynch
404	7	1.4	1274	10	09LV05	09LV05	aradidopsis	477	6	1.2	70	16	08YXQ6	08yxq6 anabaena sp
405	7	1.4	1294	5	08IQ10	08IQ10	dirosophila	478	6	1.2	71	2	09L7I0	09l7i0 synchococc
406	7	1.4	1402	5	09VC63	09VC63	dirosophila	479	6	1.2	71	17	09HMX5	09hmx5 halobacteri
407	7	1.4	1461	16	005819	005819	streptomyc	480	6	1.2	73	16	08Z521	08z521 salmonella
408	7	1.4	1553	16	08CK29	08CK29	dirosophila	481	6	1.2	73	17	08YK6	08yk6 methanopyru
409	7	1.4	1594	5	09V6W7	09V6W7	dirosophila	482	6	1.2	74	2	0931H1	0931h1 wolinnella s
410	7	1.4	1940	16	09RUC8	09RUC8	deinococcus	483	6	1.2	74	17	08T153	08t153 methanosarc
411	7	1.4	1941	16	08G751	08G751	bifidobacte	484	6	1.2	76	6	08SEP20	08sep20 macaca mula
412	7	1.4	1947	2	09XCJ2	09XCJ2	salmonella	485	6	1.2	77	2	08KMH5	08kmh5 mycoplasma
413	7	1.4	2108	13	098U19	098U19	gallus gall	486	6	1.2	77	2	08KRV5	08krv5 mycoplasma
414	7	1.4	2134	12	09YLS4	09YLS4	avian encerp	487	6	1.2	77	2	08KH20	08kh20 mycoplasma
415	7	1.4	2275	17	08TSE8	08TSE8	methanosarc	488	6	1.2	77	2	08KH21	08kh21 mycoplasma
416	7	1.4	2435	16	08ZNV6	08ZNV6	salmonella	489	6	1.2	77	2	08KH10	08kh10 mycoplasma
417	7	1.4	2467	4	015050	015050	homo sapien	490	6	1.2	77	11	08R576	08r576 mus musculu
418	7	1.4	2603	6	09XS53	09XS53	potorous tr	491	6	1.2	77	16	08X8R3	08x8r3 escherichia
419	7	1.4	2691	4	09NS12	09NS12	homo sapien	492	6	1.2	77	16	08RQW5	08rqw5 corynebacte
420	7	1.4	2692	4	09Y503	09Y503	homo sapien	493	6	1.2	77	16	08FL10	08fl10 leptospira
421	7	1.4	2705	4	095103	095103	homo sapien	494	6	1.2	77	16	08EANO	08eano shewannella
422	7	1.4	2725	4	09NYS5	09NYS5	homo sapien	495	6	1.2	77	17	028902	028902 archaeglob
423	7	1.4	3080	12	09WAC2	09WAC2	zucchini ye	496	6	1.2	78	15	078638	078638 human immun
424	7	1.4	3413	2	054593	054593	amycolatops	497	6	1.2	78	16	08ZJW7	08zjw7 salmonella
425	7	1.4	4498	5	09W223	09W223	dirosophila	498	6	1.2	78	16	08ZOV3	08zov3 salmonella
426	7	1.4	9477	2	09L4X3	09L4X3	streptomyc	499	6	1.2	79	2	093619	093619 strephyllococ
427	7	1.4	11096	2	09L4W3	09L4W3	methanobact	500	6	1.2	79	12	010321	010321 oxyria psen
428	6	1.2	17	1	09V2X3	09V2X3	homo sapien	501	6	1.2	79	15	08I150	08i150 chimpanzee
429	6	1.2	24	4	09Q6V1	09Q6V1	heparitis c	502	6	1.2	80	2	084933	084933 streptococc
430	6	1.2	26	12	09Q6V1	09Q6V1	lampetra pl	503	6	1.2	80	15	08URM2	08urm2 human immun
431	6	1.2	27	13	057557	057557	escherichia	504	6	1.2	80	15	08URM3	08urm3 human immun
432	6	1.2	28	2	047354	047354	escherichia	505	6	1.2	80	16	08X8C0	08x8c0 escherichia
433	6	1.2	30	10	08RV13	08RV13	pinus pinas	506	6	1.2	80	16	08U621	08u621 agrobacteri
434	6	1.2	31	11	08CSM5	08CSM5	mus musculu	507	6	1.2	81	4	014749	014749 homo sapien
435	6	1.2	34	16	08EBY3	08EBY3	shewannella	508	6	1.2	82	9	08SC60	08sc60 stx2 conver
436	6	1.2	36	16	08DPT5	08DPT5	vibrio vuln	509	6	1.2	82	16	098C08	098c08 rhizobium 1
437	6	1.2	42	2	053518	053518	escherichia	510	6	1.2	82	16	08Z2U2	08z2u2 salmonella
438	6	1.2	42	13	09DF18	09DF18	myxococphal	511	6	1.2	83	3	09US96	09us96 schizosacch
439	6	1.2	42	16	092MW2	092MW2	rhizobium m	512	6	1.2	83	5	08WYM9	08wym9 boletenia vi
440	6	1.2	44	2	09F526	09F526	escherichia	513	6	1.2	83	16	099U58	099u58 stephylococ
441	6	1.2	47	5	018132	018132	achacina fu	514	6	1.2	84	2	091LAH4	091lah4 thlobacillu
442	6	1.2	49	16	08EK96	08EK96	brassica na	515	6	1.2	84	15	089145	089145 chimpanzee
443	6	1.2	50	10	09AXV4	09AXV4	brassica ca	516	6	1.2	84	17	08ZUM5	08zum5 pyrobaculum
444	6	1.2	50	10	09AXV5	09AXV5	brassica na	517	6	1.2	85	15	089143	089143 chimpanzee
445	6	1.2	50	16	09PF64	09PF64	xyella faa	518	6	1.2	86	3	096UD1	096ud1 neurospora
446	6	1.2	52	10	09AXV7	09AXV7	brassica na	519	6	1.2	87	13	090452	090452 brachydanto
447	6	1.2	53	16	08PJ79	08PJ79	cinamus maj	520	6	1.2	87	15	089146	089146 chimpanzee
448	6	1.2	55	4	096F89	096F89	homo sapien	521	6	1.2	88	2	08FXR2	08fxr2 bruceella su
449	6	1.2	55	8	0958E0	0958E0	zebra mays	522	6	1.2	88	2	09KUR7	09kur7 myxococcus
450	6	1.2	55	10	08S2X3	08S2X3	zebra mays	523	6	1.2	90	5	0966J2	0966j2 caenorhabdi
451	6	1.2	55	16	09PGK5	09PGK5	xyella faa	524	6	1.2	90	6	002735	002735 sus scrofa
452	6	1.2	56	2	09L7F8	09L7F8	synchococc	525	6	1.2	90	10	091PB1	091pb1 aradidopsis
453	6	1.2	56	10	08RUB4	08RUB4	zea mays	526	6	1.2	91	7	08SNV9	08snv9 homo sapien
454	6	1.2	57	13	091566	091566	xenopus lae	527	6	1.2	91	7	030205	030205 homo sapien

528	6	1.2	91	7	019515	019515 homo sapien	601	6	1.2	108	16	099XJ9	099XJ9 streptococ
529	6	1.2	91	10	095RX6	095RX6 arabidopsis	602	6	1.2	108	16	08ZJ93	08ZJ93 yersinia pe
530	6	1.2	91	16	08F4X0	08F4X0 leptospira	603	6	1.2	108	16	08FCX0	08FCX0 escherichia
531	6	1.2	92	8	09MP63	09MP63 larix smelti	604	6	1.2	109	5	08SV06	08SV06 encephalito
532	6	1.2	92	8	09MP64	09MP64 pseudotsuga	605	6	1.2	109	5	08ITK8	08ITK8 drosoophila
533	6	1.2	92	8	09MP55	09MP55 cypae panzn	606	6	1.2	109	5	08ITJ7	08ITJ7 drosoophila
534	6	1.2	92	8	09MP66	09MP66 picea smth	607	6	1.2	109	6	09N0W5	09N0W5 oryctolagus
535	6	1.2	92	13	057377	057377 brachydanio	608	6	1.2	109	16	08P564	08P564 xanthomnas
536	6	1.2	92	15	088095	088095 chimpanzee	609	6	1.2	109	17	057992	057992 pyrococcus
537	6	1.2	92	17	026507	026507 methanobact	610	6	1.2	110	5	08IEX8	08IEX8 typanosoma
538	6	1.2	93	16	09PBR0	09PBR0 xyella fas	611	6	1.2	110	16	08ED84	08ED84 shewanella
539	6	1.2	95	4	09URF4	09URF4 homo sapien	612	6	1.2	112	2	09L482	09L482 uncultured
540	6	1.2	95	16	098LX8	098LX8 rhizobium 1	613	6	1.2	112	2	09L481	09L481 uncultured
541	6	1.2	96	4	096117	096117 homo sapien	614	6	1.2	112	2	09L476	09L476 uncultured
542	6	1.2	96	17	08ZGZ7	08ZGZ7 pyrobaculum	615	6	1.2	112	2	09L479	09L479 uncultured
543	6	1.2	97	2	049073	049073 mycoplasma	616	6	1.2	112	5	08W0J0	08W0J0 sporopetra
544	6	1.2	98	11	08R5B8	08R5B8 mus musculu	617	6	1.2	112	5	08ITK9	08ITK9 drosoophila
545	6	1.2	98	16	08FKQ0	08FKQ0 escherichia	618	6	1.2	112	16	08ZDR9	08ZDR9 yersinia pe
546	6	1.2	99	2	033447	033447 proteus mir	619	6	1.2	113	6	096934	096934 ceratilis c
547	6	1.2	99	10	08LGD1	08LGD1 arabidopsis	620	6	1.2	113	6	029367	029367 sus scrofa
548	6	1.2	99	12	08OKZ2	08OKZ2 malvastrum	621	6	1.2	113	10	09L4S7	09L4S7 oryza sativ
549	6	1.2	100	2	08VSA7	08VSA7 shigella fl	622	6	1.2	113	15	09WK54	09WK54 human immun
550	6	1.2	100	5	0963A7	0963A7 leishmania	623	6	1.2	113	15	09WK72	09WK72 human immun
551	6	1.2	100	12	09WB72	09WB72 hepatitis c	624	6	1.2	113	15	09WK52	09WK52 human immun
552	6	1.2	100	12	09WB74	09WB74 hepatitis c	625	6	1.2	113	15	09WK53	09WK53 human immun
553	6	1.2	100	12	09WB56	09WB56 hepatitis c	626	6	1.2	114	2	P71419	P71419 klebsiella
554	6	1.2	100	12	09WB39	09WB39 hepatitis c	627	6	1.2	114	16	09RUX6	09RUX6 deinococcus
555	6	1.2	100	12	09WB70	09WB70 hepatitis c	628	6	1.2	114	17	09ZBA7	09ZBA7 thermoplasm
556	6	1.2	100	12	09WB63	09WB63 hepatitis c	629	6	1.2	115	10	08L648	08L648 deschampsia
557	6	1.2	100	12	09WB71	09WB71 hepatitis c	630	6	1.2	115	11	060926	060926 mus musculu
558	6	1.2	100	12	09WB59	09WB59 hepatitis c	631	6	1.2	115	15	056909	056909 human immun
559	6	1.2	100	12	09WB54	09WB54 hepatitis c	632	6	1.2	116	9	09FEV3	09FEV3 bacterioph
560	6	1.2	100	12	09W909	09W909 hepatitis c	633	6	1.2	116	16	09YHUA	09YHUA pseudomonas
561	6	1.2	100	17	09YG98	09YG98 aeropyrum p	634	6	1.2	116	17	09YEM9	09YEM9 aeropyrum p
562	6	1.2	101	11	08BPR9	08BPR9 mus musculu	635	6	1.2	117	2	09R822	09R822 helicobacte
563	6	1.2	101	16	096Q87	096Q87 mycoplasma	636	6	1.2	117	5	027717	027717 plasmodium
564	6	1.2	101	17	09YDG4	09YDG4 aeropyrum p	637	6	1.2	117	5	08I1V1	08I1V1 plasmodium
565	6	1.2	102	12	059764	059764 schizosacch	638	6	1.2	117	10	0943X9	0943X9 oryza sativ
566	6	1.2	102	12	09J2M0	09J2M0 macaca mula	639	6	1.2	117	15	073225	073225 human immun
567	6	1.2	102	14	08BPN7	08BPN7 uncultured	640	6	1.2	117	15	073223	073223 human immun
568	6	1.2	102	14	08BPN2	08BPN2 uncultured	641	6	1.2	117	16	08DMK1	08DMK1 synchococ
569	6	1.2	102	14	08BPN7	08BPN7 uncultured	642	6	1.2	117	16	08DE56	08DE56 vibrio vuln
570	6	1.2	102	14	08BPM1	08BPM1 uncultured	643	6	1.2	118	9	08HAE8	08HAE8 salmonella
571	6	1.2	102	14	08BFM0	08BFM0 uncultured	644	6	1.2	118	16	09LOF4	09LOF4 streptomyce
572	6	1.2	102	14	08BFL9	08BFL9 uncultured	645	6	1.2	120	10	081056	081056 arabidopsi
573	6	1.2	102	14	08BFL6	08BFL6 uncultured	646	6	1.2	120	16	035601	035601 helicobacte
574	6	1.2	102	14	08BFL3	08BFL3 uncultured	647	6	1.2	120	16	09ZKQ4	09ZKQ4 helicobacte
575	6	1.2	102	14	08BFL1	08BFL1 uncultured	648	6	1.2	120	16	08FRF4	08FRF4 corynebacte
576	6	1.2	102	14	08BFL8	08BFL8 uncultured	649	6	1.2	121	15	09WM41	09WM41 human immun
577	6	1.2	102	14	08BFK7	08BFK7 uncultured	650	6	1.2	121	16	08EVJ1	08EVJ1 mycoplasma
578	6	1.2	102	14	08BFL4	08BFL4 uncultured	651	6	1.2	121	17	08PYL0	08PYL0 methanosarc
579	6	1.2	102	14	08BFL3	08BFL3 uncultured	652	6	1.2	122	2	09SOR2	09SOR2 shewanella
580	6	1.2	102	14	08BFL5	08BFL5 uncultured	653	6	1.2	122	2	09REDE	09REDE burkholderi
581	6	1.2	103	10	09XJ71	09XJ71 cucumis sat	654	6	1.2	122	16	08ZLF1	08ZLF1 salmonella
582	6	1.2	103	11	08BOK2	08BOK2 mus musculu	655	6	1.2	122	16	08DEJ8	08DEJ8 vibrio vuln
583	6	1.2	104	6	09SN86	09SN86 ovis aries	656	6	1.2	123	2	045346	045346 bacillus pa
584	6	1.2	104	16	08GCV8	08GCV8 bruceella su	657	6	1.2	123	16	09HUX0	09HUX0 pseudomonas
585	6	1.2	105	6	09GMC0	09GMC0 canis famli	658	6	1.2	124	5	09YFQ7	09YFQ7 drosoophila
586	6	1.2	105	16	08FOI5	08FOI5 leptospira	659	6	1.2	124	16	09RTY4	09RTY4 deinococcus
587	6	1.2	107	5	09APX2	09APX2 pseudomonas	660	6	1.2	124	16	0988A7	0988A7 rhizobium 1
588	6	1.2	107	5	09UZY3	09UZY3 caenorhabdi	661	6	1.2	125	5	08SWS1	08SWS1 encephalito
589	6	1.2	107	5	08ITL0	08ITL0 drosoophila	662	6	1.2	125	10	08H354	08H354 oryza sativ
590	6	1.2	107	5	08IO53	08IO53 drosoophila	663	6	1.2	125	12	09EME9	09EME9 amesacta moo
591	6	1.2	107	15	09JDX4	09JDX4 human immun	664	6	1.2	125	16	08USK7	08USK7 agrobacteri
592	6	1.2	107	15	09JDX3	09JDX3 human immun	665	6	1.2	125	17	08ULI8	08ULI8 pyrococcus
593	6	1.2	108	2	09LAI5	09LAI5 thibacilli	666	6	1.2	126	11	09CRD6	09CRD6 mus musculu
594	6	1.2	108	5	08IQ07	08IQ07 drosoophila	667	6	1.2	126	12	09WSV2	09WSV2 feline cali
595	6	1.2	108	10	08SSS9	08SSS9 oryza sativ	668	6	1.2	126	15	08URH1	08URH1 human immun
596	6	1.2	108	10	08SSS6	08SSS6 oryza sativ	669	6	1.2	126	15	08URH2	08URH2 human immun
597	6	1.2	108	10	094CY1	094CY1 oryza sativ	670	6	1.2	126	16	09PE51	09PE51 xyella fas
598	6	1.2	108	10	094HY0	094HY0 oryza sativ	671	6	1.2	126	16	09BKS5	09BKS5 bacillus ha
599	6	1.2	108	10	09AWY9	09AWY9 oryza sativ	672	6	1.2	126	16	08YV74	08YV74 anaeroba sp
600	6	1.2	108	10	08GSR6	08GSR6 oryza sativ	673	6	1.2	127	4	08NBX4	08NBX4 homo sapien

674	6	1.2	127	10	Q940E4	Q940E4 caetanea sa	747	6	1.2	141	16	Q8RFT1	Q8RFT1 fusobacteri
675	6	1.2	127	10	Q22942	Q22942 aradidopsis	748	6	1.2	141	16	Q8KG19	Q8KG19 chlorobium
676	6	1.2	127	11	Q70192	Q70192 mus musculu	749	6	1.2	142	2	Q87820	Q87820 rhizobium m
677	6	1.2	127	16	Q97HC2	Q97HC2 clostridium	750	6	1.2	142	2	Q87784	Q87784 ochrobactru
678	6	1.2	127	16	Q8PNO7	Q8PNO7 xanthomonas	751	6	1.2	142	2	Q9AF13	Q9AF13 aeromonas h
679	6	1.2	128	16	Q8VK84	Q8VK84 mycobacteri	752	6	1.2	142	2	Q87824	Q87824 rhodobacter
680	6	1.2	129	10	Q9LSH0	Q9LSH0 aradidopsis	753	6	1.2	142	4	Q9NY13	Q9NY13 homo sapien
681	6	1.2	129	10	Q9STU3	Q9STU3 aradidopsis	754	6	1.2	142	10	Q8CTB4	Q8CTB4 chlorella v
682	6	1.2	129	10	Q8LDS7	Q8LDS7 aradidopsis	755	6	1.2	142	15	Q9NM55	Q9NM55 human immun
683	6	1.2	129	17	Q9HNS4	Q9HNS4 halobacteri	756	6	1.2	142	15	Q9NM31	Q9NM31 human immun
684	6	1.2	130	2	Q8RK36	Q8RK36 pasteurella	757	6	1.2	142	15	Q9NM62	Q9NM62 human immun
685	6	1.2	130	8	Q8U034	Q8U034 exoneurella	758	6	1.2	142	15	Q9NM53	Q9NM53 human immun
686	6	1.2	130	8	Q8U033	Q8U033 exoneurella	759	6	1.2	142	15	Q9NM48	Q9NM48 human immun
687	6	1.2	130	16	Q9RI71	Q9RI71 streptomyce	760	6	1.2	142	16	Q8YHQ1	Q8YHQ1 bruceella me
688	6	1.2	130	16	Q8EBJ2	Q8EBJ2 shewanella	761	6	1.2	142	16	Q8XJH8	Q8XJH8 clostridium
689	6	1.2	130	16	Q8DM97	Q8DM97 streptococc	762	6	1.2	142	16	Q8XJH8	Q8XJH8 clostridium
690	6	1.2	130	16	Q8CMU4	Q8CMU4 streptococc	763	6	1.2	143	2	P72340	P72340 rhodobacter
691	6	1.2	131	2	P72240	P72240 pseudomonas	764	6	1.2	143	2	Q8KJF3	Q8KJF3 rhizobium e
692	6	1.2	131	5	Q23478	Q23478 caenorhabdi	765	6	1.2	143	9	Q9MCC9	Q9MCC9 lactobacill
693	6	1.2	131	8	Q35865	Q35865 strongyloce	766	6	1.2	143	11	Q9CM84	Q9CM84 mus musculu
694	6	1.2	131	8	Q79125	Q79125 brevineura	767	6	1.2	143	12	Q72518	Q72518 sindbis vir
695	6	1.2	131	10	Q8RZK5	Q8RZK5 oryza sativ	768	6	1.2	143	12	Q72519	Q72519 sindbis vir
696	6	1.2	131	15	Q90823	Q90823 human immun	769	6	1.2	143	12	Q72517	Q72517 sindbis vir
697	6	1.2	132	2	Q85755	Q85755 brochothrix	770	6	1.2	143	12	Q72520	Q72520 sindbis vir
698	6	1.2	132	2	Q8MA75	Q8MA75 buddleja as	771	6	1.2	144	2	Q9E252	Q9E252 bradyrhizob
699	6	1.2	132	10	Q65172	Q65172 mesembryant	772	6	1.2	144	5	Q9LEN1	Q9LEN1 cyanophora
700	6	1.2	132	10	Q8CTB3	Q8CTB3 chaetoceros	773	6	1.2	144	5	Q9LEN1	Q9LEN1 cyanophora
701	6	1.2	132	16	Q8XV39	Q8XV39 ralsonia s	774	6	1.2	144	16	Q910G3	Q910G3 pseudomonas
702	6	1.2	132	17	Q8ZXY3	Q8ZXY3 pyrobaculum	775	6	1.2	144	16	Q910G3	Q910G3 pseudomonas
703	6	1.2	133	10	Q8GTH5	Q8GTH5 phaeodactyl	776	6	1.2	145	2	Q9RFS1	Q9RFS1 magnetospir
704	6	1.2	133	15	Q9WLU5	Q9WLU5 human immun	777	6	1.2	145	2	Q8RPI8	Q8RPI8 uncultured
705	6	1.2	133	15	Q9WLV7	Q9WLV7 human immun	778	6	1.2	145	5	Q21324	Q21324 caenorhabdi
706	6	1.2	133	15	Q9WLM1	Q9WLM1 human immun	779	6	1.2	145	16	Q8BN79	Q8BN79 xanthomonas
707	6	1.2	133	15	Q9WLU9	Q9WLU9 human immun	780	6	1.2	146	2	Q9KM02	Q9KM02 pseudomonas
708	6	1.2	133	15	Q9WLT2	Q9WLT2 human immun	781	6	1.2	146	10	Q9M4W8	Q9M4W8 citrus sine
709	6	1.2	133	15	Q9WLV6	Q9WLV6 human immun	782	6	1.2	146	10	Q9SHX5	Q9SHX5 aradidopsis
710	6	1.2	133	15	Q9YXW1	Q9YXW1 human immun	783	6	1.2	146	10	Q94BY2	Q94BY2 aradidopsis
711	6	1.2	133	15	Q9WLU3	Q9WLU3 human immun	784	6	1.2	146	12	Q80931	Q80931 human papil
712	6	1.2	133	16	Q8XJW7	Q8XJW7 escherichia	785	6	1.2	147	2	Q93R84	Q93R84 vibrio flag
713	6	1.2	134	2	P71275	P71275 enterobacte	786	6	1.2	147	2	Q8RJ11	Q8RJ11 vibrio loge
714	6	1.2	134	10	Q944N0	Q944N0 olea europae	787	6	1.2	147	2	Q8KSW8	Q8KSW8 streptomyce
715	6	1.2	134	10	Q9AYE3	Q9AYE3 oryza sativ	788	6	1.2	147	2	Q8GCV7	Q8GCV7 citrobacter
716	6	1.2	134	10	Q93XW3	Q93XW3 brassica ca	789	6	1.2	147	6	Q8MKX9	Q8MKX9 sus scrofa
717	6	1.2	134	12	Q9QC03	Q9QC03 simian cyto	790	6	1.2	147	11	Q8K3F3	Q8K3F3 rattus norv
718	6	1.2	134	16	Q93BM1	Q93BM1 streptococc	791	6	1.2	147	16	Q97GZ1	Q97GZ1 clostridium
719	6	1.2	134	16	Q8FIAS	Q8FIAS leptospira	792	6	1.2	147	16	Q8YGO4	Q8YGO4 anabaena sp
720	6	1.2	135	15	Q9WM86	Q9WM86 human immun	793	6	1.2	147	17	Q9HMX0	Q9HMX0 halobacteri
721	6	1.2	136	16	Q8BD51	Q8BD51 vibrio vuln	794	6	1.2	148	2	Q9FAE4	Q9FAE4 alteromonas
722	6	1.2	137	5	Q9VTF3	Q9VTF3 dirosophila	795	6	1.2	148	12	Q9DMG5	Q9DMG5 rat cytomys
723	6	1.2	137	10	Q9SWY3	Q9SWY3 coriandrum	796	6	1.2	148	12	Q92YX3	Q92YX3 rhizobium m
724	6	1.2	137	10	Q8LBM4	Q8LBM4 aradidopsis	797	6	1.2	149	4	Q9BXC9	Q9BXC9 homo sapien
725	6	1.2	137	10	Q9SELE6	Q9SELE6 aradidopsis	798	6	1.2	149	11	Q9IY14	Q9IY14 mus musculu
726	6	1.2	137	15	Q9SEPO	Q9SEPO human immun	799	6	1.2	149	11	Q9CQL4	Q9CQL4 mus musculu
727	6	1.2	137	16	Q9JMT5	Q9JMT5 escherichia	800	6	1.2	149	11	Q9D5L3	Q9D5L3 mus musculu
728	6	1.2	137	16	Q8ZS54	Q8ZS54 salmonella	801	6	1.2	150	16	Q9KAY3	Q9KAY3 bacillus ha
729	6	1.2	137	16	Q8PIR6	Q8PIR6 streptococc	802	6	1.2	150	16	Q9X872	Q9X872 streptomyce
730	6	1.2	137	16	Q8P0N4	Q8P0N4 streptococc	803	6	1.2	150	16	Q8PAH7	Q8PAH7 vibrio vuln
731	6	1.2	138	4	Q9H7R1	Q9H7R1 homo sapien	804	6	1.2	151	5	Q8ISP3	Q8ISP3 branchiosto
732	6	1.2	138	10	Q8H5N8	Q8H5N8 oryza sativ	805	6	1.2	151	13	Q918E5	Q918E5 fugu rubrip
733	6	1.2	138	16	P73673	P73673 synchocyst	806	6	1.2	151	16	Q67294	Q67294 aquilex aeo
734	6	1.2	138	16	Q8E0Z3	Q8E0Z3 streptococc	807	6	1.2	152	4	Q8N124	Q8N124 homo sapien
735	6	1.2	139	2	Q44947	Q44947 borrelia bu	808	6	1.2	152	4	Q8N197	Q8N197 homo sapien
736	6	1.2	139	5	Q8SUE9	Q8SUE9 entcephalito	809	6	1.2	152	4	Q8N805	Q8N805 homo sapien
737	6	1.2	139	10	Q04652	Q04652 aradidopsis	810	6	1.2	152	12	Q9PUD0	Q9PUD0 tc virus. O
738	6	1.2	139	10	Q8G7B2	Q8G7B2 chlorella v	811	6	1.2	152	15	Q8URH4	Q8URH4 human immun
739	6	1.2	139	16	Q8U5I9	Q8U5I9 agriobacteri	812	6	1.2	152	15	Q8URH4	Q8URH4 human immun
740	6	1.2	140	2	Q9RF50	Q9RF50 magnetospir	813	6	1.2	152	16	Q97SM6	Q97SM6 streptococc
741	6	1.2	140	10	Q9FMY8	Q9FMY8 aradidopsis	814	6	1.2	152	17	Q97BJ2	Q97BJ2 thermoplaem
742	6	1.2	140	12	Q9OSP5	Q9OSP5 lymphocytic	815	6	1.2	153	2	Q82923	Q82923 escherichia
743	6	1.2	140	16	Q8G4K0	Q8G4K0 bifidobacte	816	6	1.2	153	2	Q930L4	Q930L4 escherichia
744	6	1.2	141	2	Q876B7	Q876B7 blaetobacte	817	6	1.2	153	2	Q8GEP1	Q8GEP1 pseudomonas
745	6	1.2	141	16	Q50907	Q50907 borrelia bu	818	6	1.2	153	10	Q9EFD5	Q9EFD5 aradidopsis
746	6	1.2	141	16	Q9KRV5	Q9KRV5 vibrio choli	819	6	1.2	153	11	Q9CRY8	Q9CRY8 mus musculu

820	6	1.2	153	15	Q91WN4	Q91wn4 human immun	893	6	1.2	164	4	Q9BQ52	Q9bq52 homo sapien
821	6	1.2	153	16	Q8ZK03	Q8zk03 salmonella	894	6	1.2	164	5	Q9XUV5	Q9xuv5 caenorhabdi
822	6	1.2	153	16	Q8Z0Y3	Q8z0y3 salmonella	895	6	1.2	164	10	Q8H2Q6	Q8h2q6 oryza sativ
823	6	1.2	154	5	Q8IA35	Q8ia35 brachiolesto	896	6	1.2	165	2	Q934M7	Q934m7 unclutured
824	6	1.2	154	10	Q9S724	Q9s724 arabidopsis	897	6	1.2	165	2	Q934M4	Q934m4 unclutured
825	6	1.2	154	11	Q9D217	Q9d217 mus musculu	898	6	1.2	165	2	Q8RT02	Q8rt02 unclutured
826	6	1.2	154	12	Q9J5S3	Q9j5s3 orangutan h	899	6	1.2	165	2	Q934M6	Q934m6 unclutured
827	6	1.2	154	15	Q9E8G7	Q9e8g7 human immun	900	6	1.2	165	2	Q934M5	Q934m5 unclutured
828	6	1.2	154	15	Q91WN3	Q91wn3 human immun	901	6	1.2	165	5	Q95Q19	Q95q19 caenorhabdi
829	6	1.2	154	15	Q91WN3	Q91wn3 human immun	902	6	1.2	165	6	Q9GLF4	Q9glf4 sus scrofa
830	6	1.2	155	4	Q8WMB8	Q8wmb8 homo sapien	903	6	1.2	165	10	Q65155	Q65155 arabidopsis
831	6	1.2	155	5	Q9HA73	Q9ha73 homo sapien	904	6	1.2	165	16	Q931S7	Q931s7 staphylococ
832	6	1.2	155	5	Q615Z3	Q615z3 caenorhabdi	905	6	1.2	165	16	Q8ZHJ3	Q8zhj3 yerishnia pe
833	6	1.2	155	5	Q8SSF6	Q8ssf6 encephalito	906	6	1.2	165	16	Q8XJD4	Q8xjd4 clostridium
834	6	1.2	155	15	Q91WM1	Q91wm1 human immun	907	6	1.2	165	16	Q8BDY3	Q8bdy3 vibrio vuln
835	6	1.2	155	16	Q9KXK7	Q9kxk7 streptomyce	908	6	1.2	165	17	Q9Y8U9	Q9y8u9 aeropyrum p
836	6	1.2	156	2	Q9ZAT3	Q9zat3 yerishnia en	909	6	1.2	166	10	Q9YGT6	Q9ygt6 arabidopsis
837	6	1.2	156	2	Q9SGG3	Q9sgg3 streptococc	910	6	1.2	166	10	Q8KOM0	Q8kom0 oryza sativ
838	6	1.2	156	2	Q9SGG4	Q9sgg4 streptococc	911	6	1.2	166	16	Q97GF4	Q97gf4 clostridium
839	6	1.2	156	4	Q9UJ28	Q9uj28 homo sapien	912	6	1.2	166	16	Q9X8W3	Q9x8w3 streptomyce
840	6	1.2	156	4	Q8IXF5	Q8ixf5 homo sapien	913	6	1.2	166	16	Q9XCD6	Q9xcd6 aeropyrum p
841	6	1.2	156	5	Q9U9U0	Q9u9u0 nereis vire	914	6	1.2	167	2	Q8RLT8	Q8rlt8 neisseria m
842	6	1.2	156	16	Q920F1	Q920f1 rhizobium m	915	6	1.2	167	10	Q8UD06	Q8ud06 arabidopsis
843	6	1.2	156	16	Q8PEG5	Q8peg5 xanthomonas	916	6	1.2	167	16	Q8X500	Q8x500 escherichia
844	6	1.2	157	2	Q3J3Z8	Q3j3z8 acetobacter	917	6	1.2	167	16	Q8CW26	Q8cw26 escherichia
845	6	1.2	157	5	Q8IBD9	Q8ied9 plasmodium	918	6	1.2	167	17	Q9HW22	Q9hw22 thermoplas
846	6	1.2	157	5	Q81599	Q81599 plasmodium	919	6	1.2	167	17	Q980H0	Q980h0 sulfolobus
847	6	1.2	157	11	Q9CV79	Q9cv79 mus musculu	920	6	1.2	168	2	Q8KPI9	Q8kpi9 unclutured
848	6	1.2	157	16	Q9L2C7	Q9l2c7 streptomyce	921	6	1.2	168	5	Q8ITG4	Q8itg4 plasmodium
849	6	1.2	157	16	Q8RF73	Q8rf73 corynebacte	922	6	1.2	168	16	Q8KRX5	Q8krx5 chlorobium
850	6	1.2	158	2	Q9S6B7	Q9s6b7 achromobact	923	6	1.2	168	16	Q9S2L1	Q9s2l1 streptomyce
851	6	1.2	158	4	Q9UJ27	Q9uj27 homo sapien	924	6	1.2	169	2	Q8RT01	Q8rt01 unclutured
852	6	1.2	158	10	Q8GWZ9	Q8gwz9 oryza sativ	925	6	1.2	169	2	Q69016	Q69016 zymomonas m
853	6	1.2	158	16	Q8G1C9	Q8g1c9 brucella su	926	6	1.2	169	16	Q9K3W8	Q9k3w8 streptomyce
854	6	1.2	159	2	Q8G1M5	Q8g1m5 unclutured	927	6	1.2	170	2	Q68295	Q68295 borrelia ja
855	6	1.2	159	2	Q8G1M3	Q8g1m3 unclutured	928	6	1.2	170	10	Q9STH7	Q9sth7 arabidopsis
856	6	1.2	159	2	Q8G1M2	Q8g1m2 unclutured	929	6	1.2	170	16	Q8DM61	Q8dm61 synchococc
857	6	1.2	159	2	Q8G1M1	Q8g1m1 unclutured	930	6	1.2	171	2	Q8KXJ3	Q8kxj3 unclutured
858	6	1.2	159	2	Q8G1M0	Q8g1m0 unclutured	931	6	1.2	171	8	Q8HD83	Q8hd83 panulirus j
859	6	1.2	159	2	Q8G1L9	Q8g1l9 unclutured	932	6	1.2	171	16	Q9A4Q2	Q9a4q2 caulobacter
860	6	1.2	159	2	Q8GBM1	Q8gbm1 unclutured	933	6	1.2	171	16	Q983C1	Q983c1 rhizobium 1
861	6	1.2	159	2	Q8GBM0	Q8gbm0 unclutured	934	6	1.2	172	2	Q8KPD8	Q8kpd8 unclutured
862	6	1.2	159	2	Q8GBL9	Q8gbl9 unclutured	935	6	1.2	172	2	Q8KPK2	Q8kpk2 unclutured
863	6	1.2	159	2	Q8GBL7	Q8gbl7 unclutured	936	6	1.2	172	2	Q8KPK5	Q8kpk5 unclutured
864	6	1.2	159	2	Q8GBL5	Q8gbl5 unclutured	937	6	1.2	172	2	Q8KPL8	Q8kpl8 unclutured
865	6	1.2	159	2	Q8GBL4	Q8gbl4 unclutured	938	6	1.2	172	2	Q8KPK8	Q8kpk8 unclutured
866	6	1.2	159	2	Q8GBW6	Q8gbw6 unclutured	939	6	1.2	172	2	Q8KPM4	Q8kpm4 unclutured
867	6	1.2	159	4	Q9UJZ6	Q9ujz6 homo sapien	940	6	1.2	172	2	Q8KPK1	Q8kpk1 unclutured
868	6	1.2	159	4	Q8NF74	Q8nf74 homo sapien	941	6	1.2	172	2	Q8KPK5	Q8kpk5 unclutured
869	6	1.2	159	16	Q9A5V2	Q9a5v2 caulobacter	942	6	1.2	172	2	Q8KPK4	Q8kpk4 unclutured
870	6	1.2	160	2	Q51234	Q51234 neisseria m	943	6	1.2	172	2	Q8KPM0	Q8kpm0 unclutured
871	6	1.2	160	2	Q93RF6	Q93rf6 treponema m	944	6	1.2	172	2	Q8KRL1	Q8krl1 unclutured
872	6	1.2	160	2	Q51235	Q51235 neisseria m	945	6	1.2	172	2	Q8KPM5	Q8kpm5 unclutured
873	6	1.2	160	2	Q8GN70	Q8gn70 bifidobacte	946	6	1.2	172	2	Q8KPK9	Q8kpk9 unclutured
874	6	1.2	160	11	Q8C1H8	Q8c1h8 mus musculu	947	6	1.2	172	2	Q8KPL6	Q8kpl6 unclutured
875	6	1.2	160	16	Q9X7R8	Q9x7r8 streptomyce	948	6	1.2	172	2	Q8KPL6	Q8kpl6 unclutured
876	6	1.2	161	4	Q9UK20	Q9uk20 homo sapien	949	6	1.2	172	2	Q8KPL9	Q8kpl9 unclutured
877	6	1.2	161	5	Q8SSP6	Q8ssp6 dictyosteli	950	6	1.2	172	2	Q8KPL2	Q8kpl2 unclutured
878	6	1.2	161	16	Q98CA9	Q98ca9 rhizobium 1	951	6	1.2	172	2	Q8KPL3	Q8kpl3 unclutured
879	6	1.2	161	16	Q92RWS	Q92rws rhizobium m	952	6	1.2	172	2	Q8KPK6	Q8kpk6 unclutured
880	6	1.2	161	16	Q8KA92	Q8ka92 chlorobium	953	6	1.2	172	2	Q8KPK9	Q8kpk9 unclutured
881	6	1.2	162	5	Q8SWK4	Q8swk4 encephalito	954	6	1.2	172	2	Q8KPL4	Q8kpl4 unclutured
882	6	1.2	162	10	Q8SEB2	Q8seb2 oryza sativ	955	6	1.2	172	2	Q8KPL5	Q8kpl5 unclutured
883	6	1.2	162	11	Q9CX67	Q9cx67 mus musculu	956	6	1.2	172	2	Q8KPM3	Q8kpm3 unclutured
884	6	1.2	162	16	Q9CJ98	Q9cj98 lactococcus	957	6	1.2	172	2	Q8KPM8	Q8kpm8 unclutured
885	6	1.2	162	16	Q8PEC5	Q8pec5 xanthomonas	958	6	1.2	172	2	Q9R9E8	Q9r9e8 pseudomonas
886	6	1.2	162	16	Q8EAB0	Q8eab0 shewanella	959	6	1.2	172	2	Q8K180	Q8k180 unclutured
887	6	1.2	162	17	Q8PRG4	Q8prg4 methanosarc	960	6	1.2	172	2	Q8KPK9	Q8kpk9 unclutured
888	6	1.2	163	4	Q8NNH9	Q8nnh9 homo sapien	961	6	1.2	172	2	Q8KPM1	Q8kpm1 unclutured
889	6	1.2	163	10	P93330	P93330 medicago tr	962	6	1.2	172	2	Q8KPD4	Q8kpd4 unclutured
890	6	1.2	163	10	Q64977	Q64977 arabidopsis	963	6	1.2	172	2	Q8KPM2	Q8kpm2 unclutured
891	6	1.2	163	17	Q9YAY7	Q9yay7 aeropyrum p	964	6	1.2	172	2	Q8KPK7	Q8kpk7 unclutured
892	6	1.2	164	2	Q8RT09	Q8rt09 pseudomonas	965	6	1.2	172	5	Q9T2Z8	Q9t2z8 caenorhabdi

966	6	1.2	172	10	Q8S1B6	Q8S1B6 oryza sativ
967	6	1.2	172	15	P89944	P89944 rabbit endo
968	6	1.2	172	16	Q8ZDL0	Q8ZDL0 yersinia pe
969	6	1.2	172	16	Q8CZ88	Q8CZ88 streptococ
970	6	1.2	173	2	Q8RT05	Q8RT05 uncultured
971	6	1.2	173	2	Q8RT06	Q8RT06 uncultured
972	6	1.2	173	2	Q8RT04	Q8RT04 uncultured
973	6	1.2	173	2	Q8RT03	Q8RT03 uncultured
974	6	1.2	173	12	Q8RT07	Q8RT07 uncultured
975	6	1.2	173	12	Q90B05	Q90B05 oat golden
976	6	1.2	173	16	Q8RCX6	Q8RCX6 thermoaer
977	6	1.2	173	16	Q8G4M9	Q8G4M9 bifidobacte
978	6	1.2	174	15	Q8QDK4	Q8QDK4 chimpanzee
979	6	1.2	174	15	Q8QDK3	Q8QDK3 chimpanzee
980	6	1.2	174	15	Q8QDK1	Q8QDK1 chimpanzee
981	6	1.2	174	15	Q8QDK6	Q8QDK6 chimpanzee
982	6	1.2	174	15	Q8QDK6	Q8QDK6 chimpanzee
983	6	1.2	174	17	Q8PDL6	Q8PDL6 escherichia
984	6	1.2	175	9	Q03942	Q03942 halobacteri
985	6	1.2	175	10	Q93V27	Q93V27 zea mays su
986	6	1.2	175	10	Q93W69	Q93W69 zea mays su
987	6	1.2	175	16	Q9PNW0	Q9PNW0 campylobact
988	6	1.2	175	16	Q8ZGS3	Q8ZGS3 yersinia pe
989	6	1.2	176	16	Q9Z224	Q9Z224 thizobium m
990	6	1.2	176	16	Q99RV1	Q99RV1 staphylococ
991	6	1.2	176	16	Q8CRM3	Q8CRM3 yersinia pe
992	6	1.2	177	2	Q9X690	Q9X690 mercury res
993	6	1.2	177	6	Q9GLF8	Q9GLF8 trichosurus
994	6	1.2	177	8	Q8ML25	Q8ML25 hesperantha
995	6	1.2	177	17	Q9Y911	Q9Y911 aeropyrum p
996	6	1.2	178	2	Q8RLS9	Q8RLS9 chrobactru
997	6	1.2	178	9	Q9FZ57	Q9FZ57 bacterioph
998	6	1.2	178	10	Q94LE8	Q94LE8 oryza sativ
999	6	1.2	178	16	Q9RT80	Q9RT80 delnoccuss
1000	6	1.2	178	16	Q8XNH4	Q8XNH4 clostridium

ALIGNMENTS.

RESULT 1

Q9UTB8	PRELIMINARY;	PRT;	386 AA.
AC	Q9UTB8;		
DT	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-OCT-2002 (TREMBlrel. 22, Last annotation update)		
DE	Nitrite reductase, major outer membrane copper-containing protein (EC 1.7.99.3).		
GN	AN1A OR NMA1887.		
OS	Neisseria meningitidis (serogroup A).		
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;		
OC	Neisseriaceae; Neisseria.		
OX	NCBI_TaxID=65699;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=22491 / Serogroup A / Serotype 4A;		
RX	MEDLINE=2022556; PubMed=10761919;		
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S., Jørgensen K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;		
RA	"Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."		
RT	Nature 404:502-506(2000).		
RL	EMBL; AL162757; CAB85110.1; -		
DR	HSSP; P38501; 1AS9.		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR001287; CUNO2_reductase.		
DR	Pfam; PF00394; Cu-oxidase; 2.		
DR	PRINTS; PR00695; CUNO2RDTASE.		

KM	Oxidoreductase; Complete proteome.
SO	SEQUENCE 386 AA; 40389 MW; D0624ED3A3979E88 CRC64;
Query Match	6.8%; Score 34; DB 16; Length 386;
Best Local Similarity	100.0%; Pred. No. 1e-25;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

QY	178 YHCAVAPVGMHIANGMVGLIVPEKGLPKVDKE 211
Db	169 YHCAVAPVGMHIANGMVGLIVPEKGLPKVDKE 202

RESULT 2

ID	Q9JYE1	PRELIMINARY;	PRT;	390 AA.
AC	Q9JYE1;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-OCT-2002 (TREMBlrel. 22, Last annotation update)			
DE	Major anaerobically induced outer membrane protein.			
GN	NMB1623.			
OS	Neisseria meningitidis (serogroup B).			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;			
OC	Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=491;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MC58 / Serogroup B;			
RX	MEDLINE=20175755; PubMed=10710307;			
RA	Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Cleckler A., Parksey D.S., Blair E., Citlone H., Clark E.B., Corton M.D., Uitterlind T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignan V., Pizsa M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;			
RA	"Complete genome sequence of Neisseria meningitidis serogroup B strain MC58."			
RT	Science 287:1809-1815(2000).			
RL	EMBL; AE002512; AAP41975.1; -			
DR	HSSP; P25006; INF.			
DR	TIGR; NMB1623; -			
DR	InterPro; IPR001117; Cu-oxidase.			
DR	InterPro; IPR001287; CUNO2_reductase.			
DR	Pfam; PF00394; Cu-oxidase; 2.			
DR	PRINTS; PR00695; CUNO2RDTASE.			
KW	Complete proteome.			
SO	SEQUENCE 390 AA; 40763 MW; C503F9D47DD1169D CRC64;			

Query Match	6.8%; Score 34; DB 16; Length 390;
Best Local Similarity	100.0%; Pred. No. 1.1e-25;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

QY	178 YHCAVAPVGMHIANGMVGLIVPEKGLPKVDKE 211
Db	173 YHCAVAPVGMHIANGMVGLIVPEKGLPKVDKE 206

RESULT 3

ID	Q8XPY3	PRELIMINARY;	PRT;	510 AA.
AC	Q8XPY3;			
DT	01-MAR-2002 (TREMBlrel. 20, Created)			
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)			
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)			
DE	Probable major ANAEROBICALLY induced outer membrane transmembrane protein (EC 1.7.99.3).			
GN	RSP1503 OR RSO3038.			
OS	Ralstonia solanacearum (Pseudomonas solanacearum).			
OG	Plasmid megaplasmid.			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC	Ralstoniaceae; Ralstonia.			

```

OX NCBI_TaxID=305;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlet M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RU Nature 415:497-502(2002).
DR EMBL; AL646085; CAD18654.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR001287; CuNO2 reductase.
DR InterPro; IPR000345; CytC heme_bind.
DR InterPro; IPR003088; Cyt_C1.
DR Pfam; PF00034; Cu-oxidase; 1.
DR Pfam; PF00034; Cytochrome_c; 1.
DR PRINTS; PR00695; CUNO2RDYASE.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Oxidoreductase; Plasmid; Complete proteome.
SQ SEQUENCE 510 AA; 54600 MW; 8008105DD99459AC CRC64;

Query Match 3.8%; Score 19; DB 16; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 APVGMHANGMYGLIVER 201
DB 180 APVGMHANGMYGLIVER 198
|||||
RESULT 4
ID Q8GBL8 PRELIMINARY; PRT; 159 AA.
AC Q8GBL8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Copper-containing nitrite reductase (Fragment).
GN NIRK.
OS uncultured bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=77133;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22293558; PubMed=12406765;
RA Avrahami S., Conrad R., Braker G.;
RT "Effect of Soil Ammonium Concentration on N(2)O Release and on the
RT Community Structure of Ammonia Oxidizers and Denitrifiers.";
RL Appl. Environ. Microbiol. 68:5685-5692(2002).
DR EMBL; AJ487549; CAD31803.1; -.
FT NON_TER 1 1
FT NON_TER 159 159
SQ SEQUENCE 159 AA; 17444 MW; 560E25D46ED38B4 CRC64;

Query Match 3.2%; Score 16; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKAKVE 267
DB 79 VGALTGENALKAKVE 94
|||||
RESULT 5
ID Q8BFN8 PRELIMINARY; PRT; 94 AA.
AC Q8BFN8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=A02-03-168;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
RT fragments (nirK and nirS) from nitrate and uranium-contaminated
RT groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548914; AAN59838.1; -.
FT NON_TER 1 1
FT NON_TER 94 94
SQ SEQUENCE 94 AA; 10418 MW; A79EDF8DD74B444 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 55 VGALTGENALKA 66
|||||
RESULT 6
ID Q8BFP2 PRELIMINARY; PRT; 95 AA.
AC Q8BFP2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=A07-16-44;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
RT fragments (nirK and nirS) from nitrate and uranium-contaminated
RT groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548909; AAN59833.1; -.
FT NON_TER 1 1
FT NON_TER 95 95
SQ SEQUENCE 95 AA; 10559 MW; CF4B42852051EC74 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 95;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 56 VGALTGENALKA 67
|||||
RESULT 7
ID Q8BFL7 PRELIMINARY; PRT; 99 AA.
AC Q8BFL7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS unclassified organism.
OC unclassified; environmental samples.

```

OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=001-03-134;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
fragments (nirk and nirs) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL: AF548937; AAN59861.1; -.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11019 MW; AA64C0FC7CB74CA9 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 99;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 60 VGALTGENALKA 71

RESULT 8
Q8BFN9 PRELIMINARY; PRT; 102 AA.
AC Q8BFN9;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A03-10-5;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
fragments (nirk and nirs) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL: AF548913; AAN59837.1; -.
FT NON_TER 1
FT NON_TER 102
SQ SEQUENCE 102 AA; 11327 MW; 9P317491059A3B67 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 9
Q8BFN5 PRELIMINARY; PRT; 102 AA.
AC Q8BFN5;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C01-03-6;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;

RT "Molecular diversity and characterization of nitrite reductase gene
fragments (nirk and nirs) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL: AF548917; AAN59841.1; -.
FT NON_TER 1
FT NON_TER 102
SQ SEQUENCE 102 AA; 11256 MW; 090DF6931876E087 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 10
Q8BFN3 PRELIMINARY; PRT; 102 AA.
AC Q8BFN3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B06-16-18;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
fragments (nirk and nirs) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL: AF548919; AAN59843.1; -.
FT NON_TER 1
FT NON_TER 102
SQ SEQUENCE 102 AA; 11245 MW; 915AE0E41975C548 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 11
Q8BFN0 PRELIMINARY; PRT; 102 AA.
AC Q8BFN0;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NIRK.
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B02-05-26;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrite reductase gene
fragments (nirk and nirs) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL: AF548922; AAN59846.1; -.

FT NON_TER 1 1
NON_TER 102 102
SEQUENCE 102 AA; 11276 MW; 333DF6992990FAB4 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 12

08BFM8 PRELIMINARY; PRT; 102 AA.
AC 08BFM8;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DE Nitrile reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D02-05-81;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene
fragments (nirK and nirS) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AFS48925; AAN59849.1; -.
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11286 MW; DFD696551IDCC825 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 13

08BFM2 PRELIMINARY; PRT; 102 AA.
AC 08BFM2;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Nitrile reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E01-03-40;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene
fragments (nirK and nirS) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AFS48932; AAN59856.1; -.
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11254 MW; 01AF71COPDD04069 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;

Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 14

08BF14 PRELIMINARY; PRT; 102 AA.
AC 08BF14;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Nitrile reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F02-05-104;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene
fragments (nirK and nirS) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AFS48940; AAN59864.1; -.
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11304 MW; 7972A36617960E88 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263
DB 63 VGALTGENALKA 74

RESULT 15

08BFK9 PRELIMINARY; PRT; 102 AA.
AC 08BFK9;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Nitrile reductase (Fragment).
GN NIRK.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H01-03-137;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene
fragments (nirK and nirS) from nitrate and uranium-contaminated
groundwater.";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AFS48946; AAN59869.1; -.
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11334 MW; 091839693CB90AB7 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALKA 263

Db 63 VgALTGENALKA 74

RESULT 16

Q8BF60 PRELIMINARY; PRT; 102 AA.
AC Q8BF60;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
GN Nitrile reductase (Fragment).
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DO3-10-15, E03-10-26, G03-10-86, K16-129, and K05-27;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene fragments (nitr and nirs) from nitrate and uranium-contaminated groundwater";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548924; AANS9848.1; -;
DR EMBL; AF548931; AANS9855.1; -;
DR EMBL; AF548949; AANS9872.1; -;
DR EMBL; AF548953; AANS9876.1; -;
DR EMBL; AF548956; AANS9879.1; -;
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11286 MW; 090DF6831976E087 CRC64;

Query Match 2.4%; Score 12; DB 14; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263
Db 63 VgALTGENALKA 74

RESULT 17

Q8BF60 PRELIMINARY; PRT; 103 AA.
AC Q8BF60;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Nitrile reductase (Fragment).
GN Nitr.
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A04-15-9;
RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
RT "Molecular diversity and characterization of nitrile reductase gene fragments (nitr and nirs) from nitrate and uranium-contaminated groundwater";
RL Environ. Microbiol. 0:0-0(2002).
DR EMBL; AF548912; AANS9836.1; -;
FT NON_TER 1 1
FT NON_TER 103 103
SQ SEQUENCE 103 AA; 11327 MW; 40EA68FAC498F3E CRC64;

Query Match 2.4%; Score 12; DB 14; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00094;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263
Db 64 VgALTGENALKA 75

RESULT 18

Q87751 PRELIMINARY; PRT; 142 AA.
AC Q87751;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Nitrile reductase (Fragment).
GN Nitr.
OS Hyphomicrobium zavarzinii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiaceae; Hyphomicrobium.
OX NCBI_TaxID=48292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IPAM ZV-622;
RX MEDLINE=98432797; PubMed=9758798;
RA Braker G., Feseefeldt A., Witzel K.P.;
RT "Development of PCR primer systems for the amplification of nitrile reductase genes (nirs and nitr) to detect denitrifying bacteria in environmental samples";
RL Appl. Environ. Microbiol. 64:3769-3775(1998).
DR EMBL; AJ224902; CAA12205.1; -;
DR HSSP; P25006; INIF.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1 1
FT NON_TER 142 142
SQ SEQUENCE 142 AA; 15583 MW; F76814D01D4FD1BA CRC64;

Query Match 2.4%; Score 12; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263
Db 72 VgALTGENALKA 83

RESULT 19

Q8KPL1 PRELIMINARY; PRT; 172 AA.
AC Q8KPL1;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Nitrile reductase (Fragment).
GN Nitr.
OS uncultured bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=77133;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=2194100; PubMed=11916709;
RA Prieme A., Braker G., Tiedje J.M.;
RT "Diversity of nitrile reductase (nitr and nirs) gene fragments in forested upland and wetland soils";
RL Appl. Environ. Microbiol. 68:1893-1900(2002).
DR EMBL; AY121543; AAM94074.1; -;
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 1.
FT NON_TER 1 1
FT NON_TER 172 172
SQ SEQUENCE 172 AA; 18832 MW; 8A7D4E22E814B686 CRC64;

Query Match 2.4%; Score 12; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 VgALTGENALKA 263
Db 64 VgALTGENALKA 75

Db 85 VGALTGENALKA 96

RESULT 20

Q9A007 PRELIMINARY; PRT; 360 AA.
 AC Q9A007; 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Disimilatory nitrite reductase.
 GN NIRK.
 OS Alcaligenes sp. STC1.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Alcaligenes.
 OX NCBI_TaxID=133923;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=STC1;
 RA Shoun H., Takaya N.;
 RT "Alcaligenes sp. STC1 nitrite reductase gene";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A8046603; BAB21510.1; -.
 DR HSSP: P25006; INIF.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR001287; CuNO2 reductase.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR PRINTS: PR00695; CUNO2RDTASE.
 DR PROSITE: PS00283; SOYBEAN KUNITZ; 1.
 SQ SEQUENCE 360 AA; 39072 MW; CAF063CB38430380 CRC64;

Query Match 2.4%; Score 12; DB 2; Length 360;
 Best Local Similarity 100.0%; Pred. No. 0.003;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDFAATG 148
 Db 111 MPNVDFAATG 122

RESULT 21

Q8KKH4 PRELIMINARY; PRT; 486 AA.
 AC Q8KKH4; 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Nitrite reductase precursor.
 GN NIRK.
 OS Hyphomicrobium denitrificans.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Hyphomicrobiaceae; Hyphomicrobium.
 OX NCBI_TaxID=53399;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A151;
 RA Katsuka K., Fukui A., Kabayashi M., Yamaguchi K., Suzuki S.;
 RT "Cloning and expression of copper-containing nitrite reductase from
 RT Hyphomicrobium denitrificans.";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB076606; BAC00912.1; -.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR001287; CuNO2 reductase.
 DR PRINTS: PR00695; CUNO2RDTASE.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
 KW Signal.
 FT SIGNAL.
 SQ SEQUENCE 486 AA; 52238 MW; 6984E02ECF791AF CRC64;

Query Match 2.4%; Score 12; DB 2; Length 486;
 Best Local Similarity 100.0%; Pred. No. 0.004;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 TSSFHVIGEIFD 292
 Db 400 TSSFHVIGEIFD 411

RESULT 22

Q8BFL0 PRELIMINARY; PRT; 102 AA.
 AC Q8BFL0; 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Nitrite reductase (Fragment).
 GN NIRK.
 OS uncultured organism.
 OC unclassified; environmental samples.
 OX NCBI_TaxID=155900;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H02-05-118;
 RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
 RT "Molecular diversity and characterization of nitrite reductase gene
 RT fragments (nirK and nirS) from nitrate and uranium-contaminated
 RT groundwater.";
 RL Environ. Microbiol. 0:0-0(2002).
 DR EMBL: AF548945; AAN59868.1; -.
 DR FT NON_TER 1 1
 FT NON_TER 102 102
 SQ SEQUENCE 102 AA; 11345 MW; 0909B399BC09E187 CRC64;

Query Match 2.2%; Score 11; DB 14; Length 102;
 Best Local Similarity 100.0%; Pred. No. 0.0098;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VGALTGENALK 262
 Db 63 VGALTGENALK 73

RESULT 23

Q8BPK6 PRELIMINARY; PRT; 102 AA.
 AC Q8BPK6; 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Nitrite reductase (Fragment).
 GN NIRK.
 OS uncultured organism.
 OC unclassified; environmental samples.
 OX NCBI_TaxID=155900;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G02-05-110;
 RA Fields M.W., Yan T., Wu L., Zu Y., Tiedje J.M., Zhou J.;
 RT "Molecular diversity and characterization of nitrite reductase gene
 RT fragments (nirK and nirS) from nitrate and uranium-contaminated
 RT groundwater.";
 RL Environ. Microbiol. 0:0-0(2002).
 DR EMBL: AF548950; AAN59873.1; -.
 DR FT NON_TER 1 1
 FT NON_TER 102 102
 SQ SEQUENCE 102 AA; 11264 MW; 45B54C96EEDD98D4 CRC64;

Query Match 2.2%; Score 11; DB 14; Length 102;
 Best Local Similarity 100.0%; Pred. No. 0.0098;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GALTGENALKA 263
 Db 64 GALTGENALKA 74

RESULT 24

Q9P9H9

PRELIMINARY; PRT; 361 AA.

AC Q9P9H9; 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DE Copper-containing dissimilatory nitrite reductase precursor

DE (EC 1.7.99.3).

GN NTRK.

OS Halorcula marismortui (Halobacterium marismortui).

OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

OC Halobacteriaceae; Halorcula.

OX NCBI_TaxID=2238;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21311739; PubMed=11418554;

RA Ichiki H., Tanaka Y., Mochizuki K., Yoshimatsu K., Sakurai T.,

RA Fujiwara T.;

RT "Purification, characterization, and genetic analysis of Cu-containing

RT dissimilatory nitrite reductase from a denitrifying halophilic

RT archaeon, Halorcula marismortui.";

RL J. Bacteriol. 183:4149-4156(2001).

DR EMBL; AJ278286; CAB93142.1; -.

DR HSSP; P38501; IAS8.

DR InterPro: IPR001117; Cu-oxidase.

DR Pfam: PF00394; Cu-oxidase; 1.

DR PRINTS; PR00695; CONO2REDTASE.

KW Oxidoreductase; Signal.

FT SIGNAL 1 31

FT CHAIN 32 361

Db 191 ILVPEKGLP 200

POTENTIAL.
COPPER-CONTAINING DISSIMILATORY NITRITE
REDUCTASE

SEQUENCE 361 AA; 38803 MW; 12829C586AA3F5D0 CRC64;

Query Match 2.0%; Score 10; DB 1; Length 361;

Best Local Similarity 100.0%; Pred. No. 0.34;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 ILVPEKGLP 206

Db 191 ILVPEKGLP 200

RESULT 25

Q8KPJ6

PRELIMINARY; PRT; 172 AA.

AC Q8KPJ6; 01-OCT-2002 (TReMBLrel. 22, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DE Nitrile reductase (Fragment).

GN NTRK.

OS uncultured bacterium.

OC Bacteria; environmental samples.

OX NCBI_TaxID=77133;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21914100; PubMed=11916709;

RA Pileme A., Braker G., Tiedje J.M.;

RT "Diversity of nitrile reductase (nitr and nirs) gene fragments in

RT forested upland and wetland soils.";

RL Appl. Environ. Microbiol. 68:1893-1900(2002).

DR EMBL; AY121559; AAM94090.1; -.

DR InterPro: IPR001117; Cu-oxidase.

DR Pfam; PF00394; Cu-oxidase; 1.

FT NON TER 1 1

FT NON TER 172 172

SQ SEQUENCE 172 AA; 18829 MW; 995E2A402D3DA095 CRC64;

Query Match 1.8%; Score 9; DB 2; Length 172;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 GSVGALTGE 258

Db 83 GSVGALTGE 91

RESULT 26

Q8CXB7

PRELIMINARY; PRT; 428 AA.

AC Q8CXB7; 01-MAR-2003 (TReMBLrel. 23, Created)

DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)

DE Trigger factor (Prolyl isomerase).

GN OB2078.

OS Oceanobacillus iheyensis.

OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.

OX NCBI_TaxID=182710;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=HTE831 / DSM 14371 / JCM 11309;

RA Takami H., Takaki Y., Uchiyama I.;

RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya

RT Ridge and its unexpected adaptive capabilities to extreme

RT environments.";

RL Nucleic Acids Res. 30:3927-3935(2002).

DR EMBL; AF004600; BAC14034.1; -.

KW isomerase; Complete proteome.

SQ SEQUENCE 428 AA; 48230 MW; 4D78D2CB88284C8A CRC64;

Query Match 1.8%; Score 9; DB 16; Length 428;

Best Local Similarity 100.0%; Pred. No. 4.1;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 FEFGKGNH 304

Db 179 FEFGKGNH 187

RESULT 27

Q8X8W0

PRELIMINARY; PRT; 134 AA.

AC Q8X8W0; 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

DE Regulator in colanic acid synthesis, interacts with RcsB (Regulator in

DE colanic acid synthesis RcsF).

GN RCSF OR 20208 OR ECS0198.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=O157:H7 / EDL933 / ATCC 700927;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Postel G.E., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RL Nature 409:529-533(2001).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=O157:H7 / RUMD 0509952;

RA MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubura S., Shiba T., Hattori M., Shingawa H.,
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005195; AAC54498.1; -;
DR EMBL: AP002550; BAB33621.1; -;
KM Complete proteome.
SQ SEQUENCE 134 AA; 14177 MW; 824F8209E07BB41 CRC64;

Query Match 1.6%; Score 8; DB 16; Length 134;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGCS 24
|||||
DB 10 ALMLSGCS 17

RESULT 28

Q8KPJ1 PRELIMINARY; PRT; 171 AA.
AC Q8KPJ1:
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NITRK.
OS uncultured bacterium.
OC Bacteria; environmental samples.
OK NCBI_TaxID=77133;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21914100; PubMed=11916709;
RA Prieme A., Braker G., Tiedje J.M.,
RT "Diversity of nitrite reductase (nitrk and nirs) gene fragments in
RT forested upland and wetland soils.";
RL Appl. Environ. Microbiol. 68:1893-1900(2002).
DR EMBL: AY121567; AAM94098.1; -;
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR002160; Kunitz_legume.
DR Pfam: PF00394; Cu-oxidase; 1.
DR PROSITE: PS00283; SOYBEAN_KUNITZ; 1.
FT NON_TER 1 171
FT SEQUENCE 171 AA; 18917 MW; 47E538133E7D9317 CRC64;

Query Match 1.6%; Score 8; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 VVNGSVG 253
|||||
DB 79 VVNGSVG 86

RESULT 29

Q8KPM6 PRELIMINARY; PRT; 172 AA.
AC Q8KPM6:
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Nitrite reductase (Fragment).
GN NITRK.
OS uncultured bacterium.
OC Bacteria; environmental samples.
OK NCBI_TaxID=77133;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21914100; PubMed=11916709;
RA Prieme A., Braker G., Tiedje J.M.,
RT "Diversity of nitrite reductase (nitrk and nirs) gene fragments in

RT forested upland and wetland soils.";
RL Appl. Environ. Microbiol. 68:1893-1900(2002).
DR EMBL: AY121527; AAM94058.1; -;
DR InterPro: IPR001117; Cu-oxidase.
DR Pfam: PF00394; Cu-oxidase; 1.
FT NON_TER 1 172
FT SEQUENCE 172 AA; 19069 MW; B477803BA4FCF50F CRC64;

Query Match 1.6%; Score 8; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 LTGENALK 262
|||||
DB 88 LTGENALK 95

RESULT 30

Q8TSG4 PRELIMINARY; PRT; 252 AA.
AC Q8TSG4:
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Formylmethanofuran dehydrogenase, subunit C.
GN FMDC OR MA0832.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OK NCBI_TaxID=2214;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galsagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allon N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
RA Linon L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guse A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.,
RT "The genome of *Methanosarcina acetivorans* reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010745; AAM04271.1; -;
DR InterPro: IPR002489; DUF14.
DR Pfam: PF01493; GXGKG; 1.
KM Complete proteome.
SQ SEQUENCE 252 AA; 26676 MW; 91929A25282421B1 CRC64;

Query Match 1.6%; Score 8; DB 17; Length 252;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKITVNGN 454
|||||
DB 150 GKITVNGN 157

RESULT 31

Q8PVI5 PRELIMINARY; PRT; 252 AA.
AC Q8PVI5:
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Tungsten formylmethanofuran dehydrogenase, subunit C (EC
1.2.99.5).

GN MM1978.
 OS Methanosarcina mazel (Methanosarcina frisia).
 OC Archaea: Euryarchaeota: Methanococci: Methanosarcinales;
 OC Methanosarcinaceae: Methanosarcina.
 OK NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=2120827; PubMed=12125824;
 RA Deppe-Meyer U., Johann A., Hartesch T., Merkl R., Schmitz R.A.,
 RA Martinez-Arias R., Henne A., Wieser A., Baumer S., Jacobi C.,
 RA Bruggemann H., Lienack T., Christmann A., Boemcke M., Steckel S.,
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-J., Gottschalk G.;
 RT "The genome of Methanosarcina mazel: evidence for lateral gene
 RT transfer between Bacteria and Archaea.";
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 DR EMBL: AB013435; AM31674.1; -;
 DR InterPro: IPR002489; DUF14.
 DR Pfam: PF01493; GXXGX; 1.
 SO Oxidoreductase; Complete proteome.
 KM SEQUENCE 252 AA; 26742 MW; 59CA399B11BBELAC CRC64;

Query Match 1.6%; Score 8; DB 17; Length 252;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKITVNGN 454
 DB 150 GKITVNGN 157

RESULT 32

Q9EFM7 . PRELIMINARY; PRT; 255 AA.

AC Q9EFM7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ABC transporter-like protein (Fragment).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
 OK NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=97471969; PubMed=9330910;
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT clones.";
 RL DNA Res. 4:215-230(1997).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AB005235; BAB17024.1; -;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR ATP-binding; Transport.
 KM NON_TER
 FT SEQUENCE 255 AA; 27832 MW; 79BDA0BD5378BCD5 CRC64;

Query Match 1.6%; Score 8; DB 10; Length 255;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 AAKTANAD 48
 DB 150 AAKTANAD 157

DB 129 AAKTANAD 136

RESULT 33

Q8EU09 . PRELIMINARY; PRT; 321 AA.

AC Q8EU09;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE O-sialoglycoprotein end peptidase.
 OS Mycoplasma penetrans.
 OC Bacteria: Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OK NCBI_TaxID=28227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HF-2;
 RX MEDLINE=22354719; PubMed=1246555;
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
 RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
 RT "The complete genomic sequence of Mycoplasma penetrans, an
 RT intracellular bacterial pathogen in humans.";
 RL Nucleic Acids Res. 30:5293-5300(2002).
 DR EMBL: AP004173; BAC44653.1; -;
 KM Complete proteome.
 SO SEQUENCE 321 AA; 35892 MW; EFP22FCA4DEB0BDE CRC64;

Query Match 1.6%; Score 8; DB 16; Length 321;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 IGEIPDKV 294
 DB 175 IGEIPDKV 182

RESULT 34

Q68601 . PRELIMINARY; PRT; 360 AA.

AC Q68601;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Dissimilatory copper-containing nitrite reductase.
 OS Nitr.
 OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OC Bacteria: Proteobacteria: Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Achromobacter.
 OK NCBI_TaxID=515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB11015;
 RA Suzuki E., Horikoshi N., Kohzuma T.;
 RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GF10051;
 RA Katakoka K., Furusawa H., Yamaguchi K., Suzuki S.;
 RT "Cloning and Expression of Copper Nitrite Reductase Gene from
 RT Alcaligenes xylosoxydans GF10051.";
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF051831; AAC05831.1; -;
 DR EMBL: AB013078; BAA33678.1; -;
 DR HSSP: P25006; INIF.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR001287; CuNO2_reductase.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR PRINTS: PR00695; CUNO2REDTASE.
 KM Signal.
 SO SEQUENCE 360 AA; 38939 MW; 3748B5BD3FFP44E7 CRC64;

Query Match 1.6%; Score 8; DB 2; Length 360;

RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Bell C.R., Yuan O., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
 RA Heiao J., Ziemann V., Pat G., Bowman C.L., Fujii C.Y., Vanken S.E.,
 RA Bowman C.L., Craven B., Uteback T.R., Kralak H., Feldlyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBa009381 genomic sequence."
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC024594; AAK21352.1; -
 DR HSP: Q91835.1D12.
 DR Gramene: Q9AV507; -
 DR InterPro: IPR001159; DS_RBD.
 DR InterPro: IPR002965; P_Rich_extensn.
 DR Pfam: PF00035; dsm; 2.
 DR PRINTS: PRO1217; PRICHEXTNSN.
 DR SMART: SM00358; DSRM; 2.
 DR PROSITE: PS50137; DS_RBD; 2.
 SQ SEQUENCE 514 AA; 56078 MW; A98AD36128502939 CRC64;
 Query Match 1.6%; Score 8; DB 10; Length 514;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 43 KTANADNA 50
 DB 483 KTANADNA 490
 RESULT 39
 ID Q91ZB8 PRELIMINARY; PRT; 634 AA.
 AC Q91ZB8;
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE ABC transporter-like protein (Putative ABC transporter protein).
 GN F8F6.120 OR AT5G03910.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Peters S.A., van Staveren M., Dirkee W., Stiekema W.,
 RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RT EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen W., Palm C.U., Sakurai T., Satou M., Seki M.,
 RA Shin P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
 RA Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene At5G03910 (GI:15237574).";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen W., Palm C.U.,
 RA Sakurai T., Satou M., Seki M., Shin P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;

RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AL162873; CAB85511.1; -
 DR EMBL: AY059727; AAL24084.1; -
 DR EMBL: AY091365; AAM14304.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR001140; ABC_TM_transp.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00664; ABC_membrane; 1.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 634 AA; 69197 MW; 957B5456221871A3 CRC64;
 Query Match 1.6%; Score 8; DB 10; Length 634;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 AAKTANAD 48
 DB 508 AAKTANAD 515
 RESULT 40
 ID Q93538 PRELIMINARY; PRT; 915 AA.
 AC Q93538;
 DT 01-FEB-1997 (T-EMBLrel. 02, Created)
 DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE F20D1.6 protein.
 GN F20D1.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Burton J.;
 RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z78542; CAB01748.1; -
 DR WormPep; F20D1.6; CE09499.
 SQ SEQUENCE 915 AA; 104585 MW; 0AEAF6A63A58F8566 CRC64;
 Query Match 1.6%; Score 8; DB 5; Length 915;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 87 KMETVEKV 94
 DB 503 KMETVEKV 510

Search completed: August 27, 2003, 18:45:12
 Job time : 143 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:34:22 ; Search time 42 Seconds
(without alignments)
1149.445 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502
Sequence: 1 MSKPTLIKTTLCALSLML.....NKGSLADDAVAKKTKPN 502

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : PIR_76:.*
1: p1r1:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	6.8	386	2 B81816	nitrite reductase
2	34	6.8	390	2 B81062	nitrite reductase
3	34	6.8	392	2 A49208	nitrite reductase
4	12	2.4	363	2 S32112	(EC 1.7.2.1) (EC
5	8	1.6	134	2 D64744	exopolysaccharide
6	8	1.6	134	2 F90653	regulator in colan
7	8	1.6	134	2 F85504	regulator in colan
8	8	1.6	360	2 JE0215	nitrite reductase
9	8	1.6	360	2 JG0170	nitrite reductase
10	8	1.6	389	2 D64333	pyruvate synthase
11	8	1.6	483	2 F81221	hypothetical prote
12	8	1.6	494	2 H81992	probable integral
13	8	1.6	604	2 F69802	ABC transporter (A
14	8	1.6	634	2 T48418	ABC transporter (A
15	8	1.6	915	2 T21147	hypothetical prote
16	8	1.4	86	1 CCCP55	cytochrome c555 -
17	7	1.4	98	2 AG3064	hypothetical prote
18	7	1.4	104	2 S52930	GP41 ENV protein -
19	7	1.4	109	2 S16437	ribulose-bisphosph
20	7	1.4	111	2 A70407	hypothetical prote
21	7	1.4	110	2 E86906	ribosomal protein
22	7	1.4	157	2 AF3315	transcription elon
23	7	1.4	161	2 B97802	hypothetical prote
24	7	1.4	175	2 A97742	cytochrome c limpo
25	7	1.4	191	2 AH0196	probable lipoprote
26	7	1.4	192	2 T29645	hypothetical prote
27	7	1.4	193	1 F64941	rnd protein - Bsch
28	7	1.4	193	2 C90943	probable outer mem
29	7	1.4	193	2 G85791	probable outer mem

30	7	1.4	193	2 AD0725	probable lipoprote
31	7	1.4	194	2 G69539	ribosomal protein
32	7	1.4	202	2 T36138	hypothetical prote
33	7	1.4	211	2 A85098	hypothetical prote
34	7	1.4	211	2 C84775	probable harpin-in
35	7	1.4	212	2 C70737	probable sigd prot
36	7	1.4	212	2 T29906	hypothetical prote
37	7	1.4	215	2 S55925	hypothetical arabinoga
38	7	1.4	218	2 G91207	probable replicase
39	7	1.4	220	2 T15381	hypothetical prote
40	7	1.4	238	2 I38849	LEKX-3 - human
41	7	1.4	239	2 S64327	probable membrane
42	7	1.4	245	2 B86889	zinc ABC transport
43	7	1.4	255	2 G82644	2,5-dichloro-2,5-c
44	7	1.4	258	2 AC2860	d-beta-hydroxybuty
45	7	1.4	261	2 A83196	probable transcrip
46	7	1.4	282	2 AG3047	transcription regu
47	7	1.4	282	2 AB2037	hypothetical prote
48	7	1.4	293	2 D86670	pseudouridine synt
49	7	1.4	293	2 T29907	hypothetical prote
50	7	1.4	296	2 T46617	probable chemotaxi
51	7	1.4	299	2 T52444	hypothetical prote
52	7	1.4	299	2 AC2898	glutamy-tRNA synt
53	7	1.4	299	2 D98238	helix-turn-helix d
54	7	1.4	299	2 F98285	glutamy-tRNA synt
55	7	1.4	305	2 A97637	D-beta-hydroxybuty
56	7	1.4	309	2 A81403	probable periplasm
57	7	1.4	312	2 F86044	secreted protein E
58	7	1.4	312	2 B91198	Esap protein limpo
59	7	1.4	313	2 D96616	hypothetical prote
60	7	1.4	315	2 C82933	pseudouridine synt
61	7	1.4	320	2 AD3435	cytochrome-c oxida
62	7	1.4	322	2 G97317	probable dehydroge
63	7	1.4	323	2 D87472	hypothetical prote
64	7	1.4	325	2 A86054	probable replicase
65	7	1.4	326	2 S66262	vestitone reductas
66	7	1.4	327	2 T07104	2'-hydroxydihydrod
67	7	1.4	331	2 S60671	ParB-like partitio
68	7	1.4	334	2 B69481	hypothetical prote
69	7	1.4	338	2 D97166	flagellar motor sw
70	7	1.4	341	2 A87293	hypothetical prote
71	7	1.4	341	2 A83644	hypothetical prote
72	7	1.4	352	2 B43649	avirulence protein
73	7	1.4	359	2 T22774	hypothetical prote
74	7	1.4	365	2 B48945	recombination prot
75	7	1.4	373	2 G75073	hypothetical prote
76	7	1.4	373	2 A86798	prophage p13 prote
77	7	1.4	376	2 I39582	nitrite reductase
78	7	1.4	376	2 AC3633	nitrite reductase
79	7	1.4	378	2 JG4648	nitrite reductase
80	7	1.4	378	2 E70786	hypothetical prote
81	7	1.4	379	2 T31154	hypothetical prote
82	7	1.4	387	2 B86659	Reca protein limpo
83	7	1.4	391	2 C72220	conserved hypochet
84	7	1.4	393	2 S38875	methionine adenosy
85	7	1.4	397	2 S66352	methionine adenosy
86	7	1.4	411	2 AB1095	nitrite reductase,
87	7	1.4	420	2 H75395	ABC transporter,
88	7	1.4	436	2 B55452	carriage-derived,
89	7	1.4	438	2 E83938	allantoinase BH230
90	7	1.4	438	2 C89864	coenzyme A disulfid
91	7	1.4	444	2 T03566	probable tritiger f
92	7	1.4	450	2 D36953	dihydrolipeamide S
93	7	1.4	455	2 H98191	copper-containing
94	7	1.4	459	2 S64241	hypothetical prote
95	7	1.4	462	2 AF1694	3-isopropylmalate
96	7	1.4	472	1 A49340	membrane-bound alc
97	7	1.4	495	2 A95984	probable xanthine
98	7	1.4	501	2 D87613	pilus assembly pro
99	7	1.4	502	2 D84119	ATP synthase alpha
100	7	1.4	512	2 S19354	cell fusion protei
101	7	1.4	516	2 A95134	voltage-gated chlo
102	7	1.4	521	2 A29345	steroid hormone re

probable lipoprote
ribosomal protein
hypothetical prote
hypothetical prote
probable harpin-in
probable sigd prot
hypothetical prote
hypothetical arabinoga
probable replicase
hypothetical prote
LEKX-3 - human
probable membrane
zinc ABC transport
2,5-dichloro-2,5-c
d-beta-hydroxybuty
probable transcrip
transcription regu
hypothetical prote
pseudouridine synt
hypothetical prote
probable chemotaxi
hypothetical prote
glutamy-tRNA synt
helix-turn-helix d
glutamy-tRNA synt
D-beta-hydroxybuty
probable periplasm
secreted protein E
Esap protein limpo
hypothetical prote
pseudouridine synt
cytochrome-c oxida
probable dehydroge
hypothetical prote
probable replicase
vestitone reductas
2'-hydroxydihydrod
ParB-like partitio
hypothetical prote
flagellar motor sw
hypothetical prote
hypothetical prote
avirulence protein
hypothetical prote
recombination prot
hypothetical prote
prophage p13 prote
nitrite reductase
nitrite reductase
nitrite reductase
hypothetical prote
hypothetical prote
Reca protein limpo
conserved hypochet
methionine adenosy
methionine adenosy
nitrite reductase,
ABC transporter,
carriage-derived,
allantoinase BH230
coenzyme A disulfid
probable tritiger f
dihydrolipeamide S
copper-containing
hypothetical prote
3-isopropylmalate
membrane-bound alc
probable xanthine
pilus assembly pro
ATP synthase alpha
cell fusion protei
voltage-gated chlo
steroid hormone re

103	7	1.4	521	2	E98002	conserved hypotet
104	7	1.4	525	2	T06104	2-dehydro-3-deoxy-
105	7	1.4	525	2	A41370	2-dehydro-3-deoxy-
106	7	1.4	543	2	T22585	hypothetical prote
107	7	1.4	544	2	S52081	diphosphate-fructo
108	7	1.4	550	2	T37519	probable amino aci
109	7	1.4	567	2	A71951	p-aminobenzoate sy
110	7	1.4	595	2	AH2718	aspartyl-tRNA synt
111	7	1.4	595	2	D97500	aspartyl-tRNA synt
112	7	1.4	608	2	H83284	probable very-long
113	7	1.4	611	2	F82951	glutamine-fructose
114	7	1.4	615	2	S58230	transcription init
115	7	1.4	633	2	S48956	hypothetical prote
116	7	1.4	663	2	F82160	conserved hypotet
117	7	1.4	664	2	A47012	phenol 2-monooxyge
118	7	1.4	677	2	T22534	hypothetical prote
119	7	1.4	714	2	E87543	methyalmalonyl-CoA
120	7	1.4	716	2	A75595	hypothetical prote
121	7	1.4	733	2	B40595	methyalmalonyl-CoA
122	7	1.4	744	2	IS1022	replication licens
123	7	1.4	747	2	A86659	anaerobic ribonuc
124	7	1.4	779	2	AE2402	alpha-glucosidase
125	7	1.4	807	1	IS1685	replication licens
126	7	1.4	844	2	B64678	NADH2 dehydrogenas
127	7	1.4	849	2	H71838	NADH2 dehydrogenas
128	7	1.4	859	2	S66827	aluminum resistan
129	7	1.4	923	2	T08033	serine/threonine p
130	7	1.4	957	2	E84547	probable disease r
131	7	1.4	962	2	IS3197	potassium channel
132	7	1.4	989	2	I48912	potassium channel
133	7	1.4	995	2	A88483	protein C05D11.1
134	7	1.4	1030	2	S43211	probable ATP-bindi
135	7	1.4	1046	2	T42720	cytoplasmic linker
136	7	1.4	1046	2	T42724	cytoplasmic linker
137	7	1.4	1174	1	A43736	creatine kinase (E
138	7	1.4	1174	1	A40853	potassium channel
139	7	1.4	1206	2	A64207	hypothetical prote
140	7	1.4	1214	2	G97419	streptococcal hema
141	7	1.4	1248	2	AH2637	conserved hypotet
142	7	1.4	1461	2	B70568	probable polyketid
143	7	1.4	1596	2	A33106	neurogenic locus m
144	7	1.4	1940	2	F75393	hypothetical prote
145	7	1.4	3071	2	T50345	vacuolar protein s
146	7	1.4	3413	2	T17467	rifamycin polyketi
147	6	1.2	20	2	A58903	metalloproteinase
148	6	1.2	33	1	FDF15G	antifreeze protein
149	6	1.2	33	2	A05162	antifreeze protein
150	6	1.2	42	2	A42598	urease accessory p
151	6	1.2	42	2	G95868	hypothetical prote
152	6	1.2	44	2	S52541	a 2-2 protein - in
153	6	1.2	50	2	A82759	hypothetical prote
154	6	1.2	55	2	E90630	ATP synthase Fo ch
155	6	1.2	55	2	A82823	hypothetical prote
156	6	1.2	56	2	S61509	light-harvesting p
157	6	1.2	58	2	C72388	hypothetical prote
158	6	1.2	61	1	A7KQ82	adipokinetic hormo
159	6	1.2	61	2	T13448	hypothetical prote
160	6	1.2	67	2	A82175	hypothetical prote
161	6	1.2	70	2	A11950	hypothetical prote
162	6	1.2	71	2	C84341	hypothetical prote
163	6	1.2	73	2	AC0739	probable cell divi
164	6	1.2	77	2	H69420	hydrogenase expres
165	6	1.2	77	2	H90870	hypothetical prote
166	6	1.2	77	2	A85748	unknown protein en
167	6	1.2	77	2	C64885	cell division inh1
168	6	1.2	78	2	S44993	gag polyprotein -
169	6	1.2	78	2	AG1071	conserved hypotet
170	6	1.2	79	2	T10336	hypothetical prote
171	6	1.2	80	2	H91230	hypothetical prote
172	6	1.2	80	2	G86077	hypothetical prote
173	6	1.2	80	2	S40834	hypothetical prote
174	6	1.2	80	2	AH3241	hypothetical prote
175	6	1.2	82	2	AG0946	probable CopG-fam1
176	6	1.2	83	1	CCML6	cytochrome c6 - go
177	6	1.2	83	2	F89920	conserved hypotet
178	6	1.2	85	1	CCPR6	cytochrome c6 - re
179	6	1.2	85	1	CCPR6	cytochrome c6 - pl
180	6	1.2	85	1	UCS849	cytochrome c6 - re
181	6	1.2	86	1	CCBR6	cytochrome c6 - ye
182	6	1.2	87	1	CCY66L	cytochrome c6 - sy
183	6	1.2	87	1	CCY66	cytochrome c6 - sy
184	6	1.2	88	1	CCBM6	cytochrome c6 - gr
185	6	1.2	88	1	UC7029	cytochrome c6 - gr
186	6	1.2	88	2	UCS816	cytochrome c6 [val
187	6	1.2	88	2	A83659	hypothetical prote
188	6	1.2	89	1	CCSG6	cytochrome c6 - sp
189	6	1.2	89	1	S35677	cytochrome c6 [val
190	6	1.2	89	2	S57923	cytochrome c6 - gr
191	6	1.2	90	2	E96524	protein T1N15.5 [i
192	6	1.2	91	2	D86158	F22D16.18 protein
193	6	1.2	91	2	A69153	hypothetical prote
194	6	1.2	92	2	G82603	conserved hypotet
195	6	1.2	93	2	S30186	NADH2 dehydrogenas
196	6	1.2	94	2	T06805	ribosomal protein
197	6	1.2	95	2	T17273	hypothetical prote
198	6	1.2	95	2	S77799	phosphoglycerate k
199	6	1.2	97	2	AB1666	glutamyl-tRNA (Gln)
200	6	1.2	97	2	AD1294	glutamyl-tRNA (Gln)
201	6	1.2	98	2	S70250	ferredoxin 2[Fe-4
202	6	1.2	100	2	F72751	hypothetical prote
203	6	1.2	101	2	E72691	hypothetical prote
204	6	1.2	101	2	C90564	lipoprotein [lipor
205	6	1.2	102	2	S06101	hypothetical prote
206	6	1.2	102	2	T40829	very hypotetrical
207	6	1.2	104	2	H64225	ribosomal protein
208	6	1.2	108	2	A65131	nitrite reductase
209	6	1.2	108	2	A91156	nitrite reductase
210	6	1.2	108	2	F86001	nitrite reductase
211	6	1.2	108	2	AE0020	nitrite reductase
212	6	1.2	108	2	AF1001	nitrite reductase
213	6	1.2	109	1	A55124	chlorodexin - Chlo
214	6	1.2	109	2	T02716	acidic ribosomal p
215	6	1.2	109	2	T02039	acidic ribosomal p
216	6	1.2	109	2	G71249	hypothetical prote
217	6	1.2	109	2	T02762	another specific pr
218	6	1.2	110	2	S73121	cytochrome c553 -
219	6	1.2	111	1	JO1083	cytochrome c6 prec
220	6	1.2	112	1	R5FE2E	acidic ribosomal p
221	6	1.2	112	2	S35440	acidic ribosomal p
222	6	1.2	112	2	AH0303	probable membrane
223	6	1.2	113	2	A47756	probable minor cap
224	6	1.2	114	1	R5CH2E	acidic ribosomal p
225	6	1.2	114	2	D75420	hypothetical prote
226	6	1.2	114	2	B45036	Pur beta - human (
227	6	1.2	115	2	F84931	ribonuclease p (EC
228	6	1.2	115	2	D84976	50S ribosomal prot
229	6	1.2	116	2	E83010	conserved hypotet
230	6	1.2	116	2	H71034	probable translati
231	6	1.2	116	2	E72639	hypothetical prote
232	6	1.2	117	2	C64094	ribosomal protein
233	6	1.2	117	2	UC5753	ribosomal protein
234	6	1.2	117	2	F82057	ribosomal protein
235	6	1.2	118	2	T27846	cytochrome c6 prec
236	6	1.2	120	1	A53328	cytochrome c6
237	6	1.2	120	2	T02708	hypothetical prote
238	6	1.2	120	2	T20150	hypothetical prote
239	6	1.2	120	2	B71875	hypothetical prote
240	6	1.2	120	2	C64638	hypothetical prote
241	6	1.2	123	2	A83041	conserved hypotet
242	6	1.2	124	2	A75374	hypothetical prote
243	6	1.2	125	2	E97414	hypothetical prote
244	6	1.2	126	2	B82715	50S ribosomal prot
245	6	1.2	126	2	A83890	hypothetical prote
246	6	1.2	126	2	AE2069	hypothetical prote
247	6	1.2	127	2	F97157	stage III sporulat
248	6	1.2	127	2	A84846	hypothetical prote

249	6	1.2	129	2	C83113	50S ribosomal prot	322	6	1.2	162	2	F86637	hypothetical prote
250	6	1.2	129	2	T10175	acyl carrier prote	323	6	1.2	163	2	F72565	hypothetical prote
251	6	1.2	129	2	T07643	PEARI 1 protein h	324	6	1.2	164	2	A26490	histone H1, macron
252	6	1.2	129	2	C84375	hypothetical prote	325	6	1.2	164	2	H65086	hypothetical prote
253	6	1.2	130	2	H95034	ribosomal protein	326	6	1.2	164	2	F91114	hypothetical prote
254	6	1.2	130	2	H97805	30S ribosomal prot	327	6	1.2	164	2	F85959	hypothetical prote
255	6	1.2	130	2	T37006	probable insertion	328	6	1.2	164	2	A10885	probable membrane
256	6	1.2	131	1	CCTW5T	cytochrome c552 [v	329	6	1.2	164	2	E87208	transcription elon
257	6	1.2	131	2	F64154	hypothetical prote	330	6	1.2	164	2	F70894	probable transcrip
258	6	1.2	132	2	S73552	ribosomal protein	331	6	1.2	164	2	T23249	hypothetical prote
259	6	1.2	132	2	T12229	hypothetical prote	332	6	1.2	165	2	G72486	hypothetical prote
260	6	1.2	133	2	PH1287	Ig heavy chain pre	333	6	1.2	165	2	AE0110	probable membrane
261	6	1.2	133	2	F85917	hypothetical prote	334	6	1.2	165	2	TS2114	probable transcrip
262	6	1.2	134	2	S00806	acyl carrier prote	335	6	1.2	166	2	T36616	hypothetical prote
263	6	1.2	134	2	S01257	acyl carrier prote	336	6	1.2	166	2	E97197	probable membrane
264	6	1.2	134	2	A26860	acyl carrier prote	337	6	1.2	166	2	B72607	hypothetical prote
265	6	1.2	134	2	S20499	acyl carrier prote	338	6	1.2	167	2	B64904	hypothetical prote
266	6	1.2	134	2	S10472	acyl carrier prote	339	6	1.2	167	2	D90892	fimbrial protein p
267	6	1.2	136	1	FOADH5	major core protein	340	6	1.2	167	2	E85725	probable fimbrial-
268	6	1.2	136	1	T09583	acyl carrier prote	341	6	1.2	167	2	G90175	NADH dehydrogenase
269	6	1.2	137	1	RSR128	ribosomal protein	342	6	1.2	168	2	T35737	probable integral
270	6	1.2	137	2	C84543	probable Hesh-like	343	6	1.2	169	2	T13727	conserved hypotet
271	6	1.2	137	2	AE0751	probable lipoprote	344	6	1.2	170	2	T48144	hypothetical prote
272	6	1.2	138	2	T39641	toxln 1 - Actinoba	345	6	1.2	171	2	B87593	hypothetical prote
273	6	1.2	138	2	S77160	hypothetical prote	346	6	1.2	172	2	AF0311	NADH2 dehydrogenas
274	6	1.2	139	2	T01801	acyl carrier prote	347	6	1.2	172	2	C97894	hypothetical prote
275	6	1.2	139	2	AH1847	phosphatase I reac	348	6	1.2	174	2	T08231	hypothetical prote
276	6	1.2	139	2	F97438	hypothetical prote	349	6	1.2	175	2	T13223	protein R175 - Lac
277	6	1.2	141	2	C69007	ribosomal protein	350	6	1.2	175	2	UC5912	Lyso protein - Lac
278	6	1.2	141	2	G82188	hypothetical prote	351	6	1.2	175	2	ACO148	probable exported
279	6	1.2	141	2	E70208	conserved hypotet	352	6	1.2	175	2	T02188	auxin-induced prot
280	6	1.2	142	2	E70613	probable ribosomal	353	6	1.2	175	2	H81372	probable micrococ
281	6	1.2	142	2	AC3345	LSU ribosomal prot	354	6	1.2	176	2	A33382	ubiquinol-cytochro
282	6	1.2	142	2	A82517	ribonuclease P XP2	355	6	1.2	176	2	E90031	hypothetical prote
283	6	1.2	144	1	S52140	protein-tyrosine-p	356	6	1.2	176	2	F95347	protein [imported
284	6	1.2	144	2	A83113	probable type II s	357	6	1.2	177	2	F72457	hypothetical prote
285	6	1.2	145	2	A53868	myoglobin I - Liol	358	6	1.2	178	2	E75340	conserved hypotet
286	6	1.2	145	2	T29389	hypothetical prote	359	6	1.2	179	2	E75189	hypothetical prote
287	6	1.2	146	1	HEB02	hemoglobin epsilon	360	6	1.2	179	2	B72422	hypothetical prote
288	6	1.2	146	1	HEB04	hemoglobin epsilon	361	6	1.2	181	2	T79640	MHC cell surface a
289	6	1.2	147	1	HEG12	hemoglobin epsilon	362	6	1.2	181	2	C71662	hypocellulose prote
290	6	1.2	147	1	F75182	ribosomal protein	363	6	1.2	181	2	H87384	acetyltransferase,
291	6	1.2	147	2	S18713	ribosomal protein	364	6	1.2	182	2	TS1128	isopentenyl diphos
292	6	1.2	147	2	B97174	chemotaxis protein	365	6	1.2	183	2	A48983	NADH2 dehydrogenas
293	6	1.2	147	2	AF1905	hypothetical prote	366	6	1.2	183	2	T46943	mammopne synthesi
294	6	1.2	147	2	H84341	hypothetical prote	367	6	1.2	184	2	S74232	bidirectional hyd
295	6	1.2	148	1	CKM6R	cytochrome c6 prec	368	6	1.2	184	2	D70362	ribosome recycling
296	6	1.2	148	2	A55712	hypothetical prote	369	6	1.2	185	2	T08200	probable disease r
297	6	1.2	148	2	E95384	protein [imported	370	6	1.2	185	2	S59560	histone H1.41 - Ga
298	6	1.2	149	2	A47054	ferric uptake regu	371	6	1.2	185	2	D75635	resolvasase - Delnoc
299	6	1.2	149	2	A36870	ferric uptake regu	372	6	1.2	186	2	S04671	H+-transporting tw
300	6	1.2	150	2	S36292	T-cell receptor ga	373	6	1.2	186	2	B32806	resolvasase - Becher
301	6	1.2	150	2	D71031	probable ribosomal	374	6	1.2	186	2	S09632	tnpr protein - Pse
302	6	1.2	150	2	A42282	ferric uptake regu	375	6	1.2	186	4	S58115	resolvasase (EC 6.5.
303	6	1.2	150	2	A82118	ferric uptake regu	376	6	1.2	187	2	T10073	amine dehydrogenas
304	6	1.2	150	2	T36099	hypothetical prote	377	6	1.2	187	2	C83627	alkyl hydroperoxid
305	6	1.2	150	2	E42528	B2AR protein - vac	378	6	1.2	188	2	B97875	conserved hypotet
306	6	1.2	150	2	S22208	hypothetical prote	379	6	1.2	189	2	AE9024	conserved hypotet
307	6	1.2	150	2	A83919	hypothetical prote	380	6	1.2	189	2	F64533	co-chaperone and h
308	6	1.2	151	2	C70408	hypothetical prote	381	6	1.2	189	2	G72407	hypothetical prote
309	6	1.2	152	2	A95023	hypothetical prote	382	6	1.2	189	2	B95219	conserved hypotet
310	6	1.2	153	2	T00303	nuclease homolog -	383	6	1.2	189	2	H98082	conserved hypotet
311	6	1.2	153	2	S10665	hypothetical prote	384	6	1.2	190	2	B86871	hypothetical prote
312	6	1.2	153	2	AF1066	probable exported	385	6	1.2	191	2	H71973	24kDa chaperone -
313	6	1.2	155	2	T33008	hypothetical prote	386	6	1.2	193	2	H97732	calf thymus ribonu
314	6	1.2	156	2	B70077	hypothetical prote	387	6	1.2	194	2	G81437	endopeptidase Clp
315	6	1.2	158	2	AC3398	SSU ribosomal prot	388	6	1.2	195	2	S44788	D2007.2 protein -
316	6	1.2	159	2	H87539	conserved hypotet	389	6	1.2	195	1	HSXLSA	histone H5A - Afri
317	6	1.2	160	2	S74318	fimbrial protein c	390	6	1.2	196	1	ZZZRAM	nodulation protein
318	6	1.2	160	2	S74317	fimbrial protein c	391	6	1.2	196	1	ZZZRA4	nodulation protein
319	6	1.2	160	2	T35326	hypothetical prote	392	6	1.2	196	2	A30484	histone H5B - Afri
320	6	1.2	162	2	D83073	conserved hypotet	393	6	1.2	196	2	C95321	NodH N-acetyltransfe
321	6	1.2	162	2	H70779	hypothetical prote	394	6	1.2	196	2	H87279	SCO1/2 family prot

395	6	1.2	197	2	AG3237	conserved hypochet	468	6	1.2	213	2	B97947	hypothetical prote
396	6	1.2	197	2	G82973	transcription regu	469	6	1.2	214	2	B87424	outer membrane pro
397	6	1.2	197	2	T38235	hypothetical prote	470	6	1.2	215	2	AH3389	transporter BME11
398	6	1.2	198	1	FOAD72	major core protein	471	6	1.2	215	2	T45229	probable F420H2-de
399	6	1.2	198	2	E86183	hypothetical prote	472	6	1.2	216	2	T12742	hypothetical prote
400	6	1.2	198	2	AC1880	hypothetical prote	473	6	1.2	216	2	A23055	hypothetical prote
401	6	1.2	199	1	S47726	outer-membrane lip	474	6	1.2	218	2	AB3314	histone H1.01 - ch
402	6	1.2	199	1	I51280	atroxase (EC 3.4.2	475	6	1.2	218	2	AH32149	hydroxyacylglycath
403	6	1.2	199	2	A91176	outer membrane pro	476	6	1.2	219	2	B75440	hypothetical prote
404	6	1.2	199	2	B86022	outer membrane pro	477	6	1.2	219	2	T38159	rfbJ protein/conse
405	6	1.2	199	2	C71979	urease accessory p	478	6	1.2	220	2	AH1344	probable RNA-bind
406	6	1.2	199	2	D64528	urease accessory p	479	6	1.2	220	2	E72204	calf thymus ribonu
407	6	1.2	199	2	T36905	hypothetical prote	480	6	1.2	223	1	F64404	hypothetical prote
408	6	1.2	200	1	A39741	cytochrome c bioge	481	6	1.2	223	2	UC6323	hypothetical prote
409	6	1.2	200	2	AE1898	urease accessory p	482	6	1.2	223	2	T21437	endo nuclease (EC 3
410	6	1.2	200	2	A96664	unknown protein T1	483	6	1.2	223	2	A40866	hypothetical prote
411	6	1.2	201	2	B87653	hypothetical prote	484	6	1.2	223	2	T306440	H1-60-induced diff
412	6	1.2	202	2	D81378	probable integral	485	6	1.2	223	2	E75390	hypothetical prote
413	6	1.2	202	2	A82160	hypothetical prote	486	6	1.2	225	2	F87591	phage shock protei
414	6	1.2	202	2	S38922	hypothetical prote	487	6	1.2	225	2	E64075	DNA-binding respon
415	6	1.2	203	2	S46443	adamalysin (EC 3.4	488	6	1.2	225	2	*A83877	urease accessory p
416	6	1.2	203	2	AB2871	urease accessory p	489	6	1.2	226	1	B64181	hypothetical prote
417	6	1.2	203	2	S73281	hypothetical prote	490	6	1.2	227	2	T11327	nrfB protein - Hae
418	6	1.2	203	2	JC4871	phospholipase C (B	491	6	1.2	227	2	A72746	cytochrome-c oxida
419	6	1.2	204	2	AD5022	CRK-1 - human	492	6	1.2	227	2	A72746	probable ribonucle
420	6	1.2	204	2	AD3411	transporter, lyse	493	6	1.2	228	2	G97308	amino acid ABC-tyr
421	6	1.2	204	2	E72695	adenylate kinase (494	6	1.2	228	2	T38622	ribulose-phosphate
422	6	1.2	204	2	G83034	urease accessory p	495	6	1.2	228	2	AD0693	pyridoxal kinase (
423	6	1.2	204	2	F36950	urease accessory p	496	6	1.2	228	2	A98166	hypothetical prote
424	6	1.2	204	2	H90027	urease accessory p	497	6	1.2	228	2	AF3121	hypothetical prote
425	6	1.2	204	2	D97647	urease accessory p	498	6	1.2	228	2	H97336	uncharacterized co
426	6	1.2	205	2	AG3014	ureg protein (AB00	499	6	1.2	229	2	I51227	histone H1A - Afri
427	6	1.2	205	2	H98269	Holliday junction	500	6	1.2	229	2	C82165	conserved hypochet
428	6	1.2	205	2	G90794	ruva, holliday jun	501	6	1.2	229	2	C84248	galactosyltransfer
429	6	1.2	205	2	T50714	urease accessory p	502	6	1.2	230	2	E72276	phosphoribosylamin
430	6	1.2	205	2	F36138	urease accessory p	503	6	1.2	232	1	TVFV10	transforming prote
431	6	1.2	205	2	UN0755	urease accessory p	504	6	1.2	232	2	DE0163	myelin expression
432	6	1.2	205	2	A85604	urease accessory p	505	6	1.2	232	2	D90254	conserved hypochet
433	6	1.2	206	2	G82893	probable urease ac	506	6	1.2	233	2	JC7689	ubiquitin carboxyl
434	6	1.2	206	2	S74933	urease complex com	507	6	1.2	233	2	F64482	hypochelical prote
435	6	1.2	206	2	AC3478	urease accessory p	508	6	1.2	233	2	F87216	probable Tetr-fam1
436	6	1.2	207	2	A56190	protein ybJ precu	509	6	1.2	233	2	F72405	ABC transporter, A
437	6	1.2	207	2	E69085	titin - rat (fragm	510	6	1.2	234	2	B96975	probable membrane
438	6	1.2	207	2	S54128	conserved hypochet	511	6	1.2	234	2	D70829	probable regulator
439	6	1.2	208	1	BVECFB	hypochelical 21.4K	512	6	1.2	234	2	C75368	conserved hypochet
440	6	1.2	208	1	A83682	incB protein - pha	513	6	1.2	235	2	B95889	probable transcrip
441	6	1.2	208	2	AC3458	urease accessory p	514	6	1.2	235	2	G75324	hypochelical prote
442	6	1.2	208	2	T45272	urease accessory p	515	6	1.2	236	2	F83083	hypothetical prote
443	6	1.2	209	1	HSXLIa	methyltransferase	516	6	1.2	236	2	B69090	hypothetical prote
444	6	1.2	209	2	S76990	histone H1A - Afri	517	6	1.2	237	1	S56137	conserved hypochet
445	6	1.2	209	2	AD2073	hypochelical prote	518	6	1.2	237	2	F71563	membrane-bound tet
446	6	1.2	209	2	T40099	two-component resp	519	6	1.2	237	2	A81716	probable sam-depen
447	6	1.2	209	2	G82359	probable synaptoch	520	6	1.2	237	2	AB2623	conserved hypochet
448	6	1.2	210	2	T27471	conserved hypochet	521	6	1.2	237	2	AC1953	hydrolyase (impo
449	6	1.2	211	2	G64136	hypochelical prote	522	6	1.2	238	2	B72320	rubrerythrin (impo
450	6	1.2	211	2	C68309	probable DNA-tapur	523	6	1.2	238	2	S70468	ribonuclease HII -
451	6	1.2	212	1	PMBYD	uridine kinase udk	524	6	1.2	238	2	A55324	agglutinin (WBA I)
452	6	1.2	212	2	A28470	H+-transpoting tw	525	6	1.2	238	2	D86538	agglutinin (WBA I)
453	6	1.2	212	2	F64940	histone H1 - mouse	526	6	1.2	238	2	C72086	SAM-dependent meth
454	6	1.2	212	2	C90942	hypochelical prote	527	6	1.2	239	2	*G82383	oxidoreductase, sh
455	6	1.2	212	2	G85790	hypochelical prote	528	6	1.2	239	2	UC7768	blue fluorescent p
456	6	1.2	212	2	AD0714	hypochelical prote	529	6	1.2	239	2	A46243	epidermal growth f
457	6	1.2	212	2	A84349	RNAase H (imported	530	6	1.2	239	2	AD0086	fimbrial chaperone
458	6	1.2	212	2	S25996	hypochelical prote	531	6	1.2	239	2	B69305	succinoglycan bios
459	6	1.2	213	1	HSRB13	histone H1.3 - rab	532	6	1.2	241	2	H11844	Kora protein - Str
460	6	1.2	213	2	S01766	GTP-binding protei	533	6	1.2	241	2	F72385	conserved hypochet
461	6	1.2	213	2	F36364	GTP-binding protei	534	6	1.2	241	2	B87337	ada regulatory pro
462	6	1.2	213	2	S58279	ABC-binding protei	535	6	1.2	242	2	AB0165	arginine transport
463	6	1.2	213	2	D69483	hypochelical prote	536	6	1.2	242	2	AH3521	high-affinity bran
464	6	1.2	213	2	F42365	hypochelical prote	537	6	1.2	243	2	C69473	methylthiodenosin
465	6	1.2	213	2	G95079	ABC transporter, A	538	6	1.2	244	2	AB3524	transcription regu
466	6	1.2	213	2	E90396	conserved hypochet	539	6	1.2	245	2	B83462	probable short-cha
467	6	1.2	213	2	E83180	probable transcrip	540	6	1.2	246	2	C48492	3-deoxy-manno-occu
										246	2	S46604	2-deoxyglucose-6-p

541	6	1.2	246	2	S46747	614	6	1.2	269	2	B71120	hypotheical prote
542	6	1.2	247	2	S77280	615	6	1.2	269	2	AD2833	Gdfer family prote
543	6	1.2	247	2	S72859	616	6	1.2	270	2	T48780	hypotheical prote
544	6	1.2	247	2	G64788	617	6	1.2	270	2	B32835	hypotheical ntra
545	6	1.2	248	1	J01682	618	6	1.2	273	2	AG1470	conserved hypothe
546	6	1.2	248	2	AG2969	619	6	1.2	274	2	J00686	nosinepride resist
547	6	1.2	248	2	G69111	620	6	1.2	274	2	D91178	probable ARAC-type
548	6	1.2	248	2	G97340	621	6	1.2	274	2	E86024	probable ARAC-type
549	6	1.2	249	2	E71649	622	6	1.2	274	2	S47736	probable transcrip
550	6	1.2	250	2	AG2635	623	6	1.2	274	2	AB1821	hypotheical prote
551	6	1.2	250	2	F97417	624	6	1.2	275	2	D98313	glutamine ABC tran
552	6	1.2	250	2	E82880	625	6	1.2	275	2	C84673	hypotheical prote
553	6	1.2	250	2	T08680	626	6	1.2	276	2	T45275	oxidoreductase of
554	6	1.2	251	2	C83053	627	6	1.2	276	2	T44386	ribosomal protein
555	6	1.2	251	2	C75521	628	6	1.2	276	2	B25345	tropoin T, cardia
556	6	1.2	252	2	A69096	629	6	1.2	276	2	A25345	tropoin T, cardia
557	6	1.2	252	2	B83750	630	6	1.2	276	2	T01177	hypotheical prote
558	6	1.2	253	1	PMHUTM	631	6	1.2	277	2	G75520	hypotheical prote
559	6	1.2	253	1	VHVUTV	632	6	1.2	278	2	B71896	probable outer mem
560	6	1.2	253	2	JM0040	633	6	1.2	278	2	H87663	hypotheical prote
561	6	1.2	253	2	AE2585	634	6	1.2	279	2	AG0421	sugar transport sy
562	6	1.2	253	2	F69134	635	6	1.2	279	2	G81054	cytochrome c5 NMBI
563	6	1.2	253	2	C82248	636	6	1.2	279	2	F81821	probable cytochrom
564	6	1.2	254	2	E71203	637	6	1.2	279	2	B97405	hydrolase, probabl
565	6	1.2	254	2	S10929	638	6	1.2	280	2	B87547	transcription regu
566	6	1.2	254	2	A49895	639	6	1.2	281	2	F82832	pantoate-beta-alan
567	6	1.2	254	2	D69140	640	6	1.2	281	2	A12156	S-formylglutathion
568	6	1.2	254	2	C75540	641	6	1.2	282	2	AB7610	nicotinate-nucleot
569	6	1.2	255	2	D75415	642	6	1.2	282	2	B83453	hypotheical prote
570	6	1.2	255	2	A69339	643	6	1.2	282	2	F82851	conserved hypothe
571	6	1.2	255	2	AF3557	644	6	1.2	282	2	A97315	xylanase/chitin de
572	6	1.2	256	2	S14518	645	6	1.2	283	2	F87323	hypotheical prote
573	6	1.2	256	2	E72454	646	6	1.2	283	2	T12062	xpsr protein - Xan
574	6	1.2	256	2	AH1978	647	6	1.2	283	2	T13879	matunase-like prot
575	6	1.2	257	2	T01342	648	6	1.2	283	2	H97610	hypotheical prote
576	6	1.2	257	2	E90763	649	6	1.2	284	1	S71851	heat shock transcr
577	6	1.2	257	2	F85626	650	6	1.2	284	2	H85436	heat shock transcr
578	6	1.2	258	2	E86786	651	6	1.2	284	2	G64753	yagm protein - Esc
579	6	1.2	258	2	A81448	652	6	1.2	284	2	T14167	ABC transport prot
580	6	1.2	258	2	G75372	653	6	1.2	284	2	H97991	hypotheical prote
581	6	1.2	258	2	T11582	654	6	1.2	285	2	D97367	moeb-like protein
582	6	1.2	259	2	A64509	655	6	1.2	285	2	UC4315	steroidogenic acut
583	6	1.2	259	2	A44988	656	6	1.2	285	2	G72414	hypotheical prote
584	6	1.2	259	2	H95133	657	6	1.2	285	2	T29804	hypotheical prote
585	6	1.2	259	2	B82348	658	6	1.2	286	2	A10288	pyridoxal kinase (
586	6	1.2	259	2	B96954	659	6	1.2	286	2	D71565	hypotheical prote
587	6	1.2	259	2	D98002	660	6	1.2	287	2	F64320	probable pyridoxal
588	6	1.2	259	2	E86288	661	6	1.2	287	2	A90922	pyridoxal kinase 2
589	6	1.2	260	2	B82927	662	6	1.2	287	2	E85770	pyridoxal kinase 2
590	6	1.2	260	2	F71438	663	6	1.2	288	2	E64151	probable pyridoxal
591	6	1.2	260	2	F95899	664	6	1.2	288	2	S69661	hypotheical prote
592	6	1.2	261	2	E83186	665	6	1.2	288	2	C82956	pyridoxamine kinas
593	6	1.2	262	2	T22130	666	6	1.2	288	2	PH1917	FL-160.1 protein -
594	6	1.2	262	2	AG0395	667	6	1.2	289	2	C84173	heat shock protein
595	6	1.2	264	1	QRBOLD	668	6	1.2	290	2	F90352	hypotheical prote
596	6	1.2	264	1	T14261	669	6	1.2	290	2	C96776	hypotheical prote
597	6	1.2	264	2	C87605	670	6	1.2	291	2	B95316	probable ABC trans
598	6	1.2	265	2	F86628	671	6	1.2	291	2	AB1696	stationary-phase s
599	6	1.2	265	2	T36386	672	6	1.2	291	2	T20942	hypotheical prote
600	6	1.2	266	2	T37878	673	6	1.2	291	2	A98229	hypotheical membr
601	6	1.2	266	2	AC2383	674	6	1.2	291	2	H86075	hypotheical prote
602	6	1.2	267	1	TSPSNA	675	6	1.2	292	2	F71819	probable proteinas
603	6	1.2	267	1	JG6326	676	6	1.2	292	2	T09030	hypotheical prote
604	6	1.2	267	2	H96558	677	6	1.2	292	2	T35546	hypotheical prote
605	6	1.2	267	2	C72415	678	6	1.2	293	2	T04951	hypotheical prote
606	6	1.2	267	2	E83858	679	6	1.2	293	2	C90139	ribokinase (rbsk-1
607	6	1.2	268	1	A69000	680	6	1.2	293	2	E82582	DnaI protein XP223
608	6	1.2	268	1	G83640	681	6	1.2	293	2	A71543	probable acid phos
609	6	1.2	269	1	J02127	682	6	1.2	293	2	H70625	hypotheical prote
610	6	1.2	269	2	A30768	683	6	1.2	293	2	E81661	conserved hypothe
611	6	1.2	269	2	A46506	684	6	1.2	294	2	AE1653	translation elonga
612	6	1.2	269	2	S28369	685	6	1.2	294	2	A11281	translacion elonga
613	6	1.2	269	2	H69768	686	6	1.2	294	2	T35191	probable DNA hydro

687	6	1.2	236	2	A36366	enhancer-binding p	760	6	1.2	315	2	D84139	mannose-6-phosphat
688	6	1.2	236	2	F82077	nucleonate-nucleot	761	6	1.2	315	2	H85722	hypothetical prote
689	6	1.2	236	2	T51336	nucleonate-nucleot	762	6	1.2	315	2	AC3522	high-affinity bran
690	6	1.2	236	2	A40593	transcription acti	763	6	1.2	315	2	E64906	probable membrane
691	6	1.2	236	2	T46828	transcription regu	764	6	1.2	316	2	D70451	conserved hypochet
692	6	1.2	237	2	H82996	heat shock protein	765	6	1.2	316	2	D41830	hypothetical prote
693	6	1.2	237	2	E64733	nucleonate-nucleot	766	6	1.2	316	2	AG2977	hypothetical prote
694	6	1.2	237	2	A99643	quinolinate phosph	767	6	1.2	316	2	C72615	probable transport
695	6	1.2	237	2	A85494	quinolinate phosph	768	6	1.2	316	2	B70768	hypothetical prote
696	6	1.2	237	2	AH0530	nucleonate-nucleot	769	6	1.2	317	2	B82637	bifunctional biotri
697	6	1.2	238	2	B45470	hydroxymethylgluta	770	6	1.2	317	2	T38935	hypothetical prote
698	6	1.2	238	2	S75972	hypothetical prote	771	6	1.2	317	2	T10300	dUTP pyrophosphata
699	6	1.2	238	2	F84224	hypothetical prote	772	6	1.2	317	2	S63358	hypothetical prote
700	6	1.2	238	2	G90529	heat shock protein	773	6	1.2	318	2	C84651	hypothetical prote
701	6	1.2	238	2	A82879	conserved hypochet	774	6	1.2	319	2	F84966	thiooxidoxin-diulf
702	6	1.2	239	2	E69835	ribitol dehydrogen	775	6	1.2	319	2	E98305	hypothetical prote
703	6	1.2	239	2	H89906	hypothetical prote	776	6	1.2	319	2	T43940	hypothetical prote
704	6	1.2	300	2	S07540	site-specific DNA-	777	6	1.2	320	2	C81941	probable kinase/ph
705	6	1.2	300	2	S35549	site-specific DNA-	778	6	1.2	320	2	A81165	hypothetical prote
706	6	1.2	300	2	D72405	hypothetical prote	779	6	1.2	320	2	A97279	hypothetical prote
707	6	1.2	301	1	RGE0CT	GTP-binding protei	780	6	1.2	321	2	E87715	quinone oxidoreduc
708	6	1.2	301	1	J01927	nonstructural prot	781	6	1.2	321	2	AE3646	4-hydroxybutyrate
709	6	1.2	301	2	E81024	glycine-tRNA ligas	782	6	1.2	321	2	T08462	hypothetical prote
710	6	1.2	301	2	H91057	GTP-binding protei	783	6	1.2	322	2	T14597	proteasome homolog
711	6	1.2	301	2	E85902	GTP-binding protei	784	6	1.2	322	2	AD3134	hypothetical prote
712	6	1.2	301	2	AB0829	GTP-binding protei	785	6	1.2	323	2	A99211	hypothetical prote
713	6	1.2	301	2	F90115	beta subunit of G	786	6	1.2	323	2	A72531	hypothetical prote
714	6	1.2	302	1	JN0794	phosphate butyryl	787	6	1.2	324	2	A97493	hypothetical prote
715	6	1.2	302	2	C64103	glycine-tRNA ligas	788	6	1.2	324	2	AH2710	conserved hypochet
716	6	1.2	302	2	F64042	GTP-binding protei	789	6	1.2	324	2	D87075	probable DNA-bind
717	6	1.2	302	2	S72542	NSM protein - toma	790	6	1.2	325	1	H64056	protein-export pro
718	6	1.2	302	2	S52584	crte protein - Erw	791	6	1.2	325	2	A12252	hydrogenase precurs
719	6	1.2	302	2	A37802	crte protein - Erw	792	6	1.2	326	2	F75178	probable hydrogena
720	6	1.2	302	2	AB0416	nucleonate-nucleot	793	6	1.2	326	2	E71035	hypothetical prote
721	6	1.2	302	2	D85025	nonstructural prote	794	6	1.2	326	2	F72639	branched-chain alp
722	6	1.2	303	1	A42544	probable GTP-bind	795	6	1.2	327	2	B83995	hypothetical prote
723	6	1.2	303	2	AG0331	CRKL protein - hum	796	6	1.2	327	2	S61982	transporter Atu186
724	6	1.2	303	2	S41754	SH2/SH3 adaptor pr	797	6	1.2	327	2	AC2805	hypothetical prote
725	6	1.2	303	2	S58352	methionyl-tRNA for	798	6	1.2	327	2	C97584	conserved hypochet
726	6	1.2	303	2	G97734	hypothetical prote	799	6	1.2	327	2	A82335	ferric ion ABC tra
727	6	1.2	303	2	F84401	CKK-II - human	800	6	1.2	327	2	A83848	hypothetical prote
728	6	1.2	304	2	A45022	C-Crk - mouse	801	6	1.2	328	2	D83312	RING-H2 finger pro
729	6	1.2	304	2	I58394	hypothetical prote	802	6	1.2	328	2	T00747	probable pantothen
730	6	1.2	304	2	E70698	hypothetical 32.6	803	6	1.2	329	2	T35567	hypothetical prote
731	6	1.2	305	2	B64981	probable transport	804	6	1.2	329	2	F91290	hypothetical prote
732	6	1.2	305	2	B91006	probable periplasm	805	6	1.2	329	2	A86132	hypothetical prote
733	6	1.2	305	2	C85850	probable periplasm	806	6	1.2	329	2	B75615	pyridoxamine kinas
734	6	1.2	305	2	AC0778	hypothetical prote	807	6	1.2	329	2	H70744	hypothetical prote
735	6	1.2	306	2	D84039	forminoglutamate	808	6	1.2	329	2	T32272	nucleonate-nucleot
736	6	1.2	306	2	F84276	conserved hypochet	809	6	1.2	330	2	F97587	photosystem II oxy
737	6	1.2	306	2	D69298	probable esterase/	810	6	1.2	331	2	T08403	chlorodoxin reduct
738	6	1.2	306	2	D70681	hypothetical prote	811	6	1.2	332	1	B70015	signal peptidase I
739	6	1.2	306	2	T33487	gravin - human (fr	812	6	1.2	332	2	AE0331	UDP-glucose 4-epim
740	6	1.2	306	2	A43922	methionyl-tRNA for	813	6	1.2	332	2	E69148	endospore developm
741	6	1.2	308	2	H82009	methionyl-tRNA for	814	6	1.2	332	2	C48835	anexin homolog -
742	6	1.2	308	2	F81238	hypothetical prote	815	6	1.2	332	2	C40646	peptide ABC transp
743	6	1.2	308	2	T33676	ribokinase (import	816	6	1.2	332	2	T18527	ketol-acid reducto
744	6	1.2	309	2	AB0952	hypothetical prote	817	6	1.2	333	2	H87120	ketol-acid reducto
745	6	1.2	309	2	AE3309	probable membrane	818	6	1.2	333	2	D70855	probable replicati
746	6	1.2	310	2	A55050	endothelial monocy	819	6	1.2	333	2	B96657	probable low-affin
747	6	1.2	310	2	S69053	beta-lactamase (EC	820	6	1.2	333	2	H75458	hypothetical prote
748	6	1.2	311	1	S02714	hypothetical prote	821	6	1.2	333	2	T36397	sugar phosphate tr
749	6	1.2	311	2	T41284	hypothetical prote	822	6	1.2	333	2	JX0343	tricyclicglycerol 11
750	6	1.2	311	2	B84537	conserved hypochet	823	6	1.2	333	2	S15941	PEP-fructosephosph
751	6	1.2	312	2	B69303	hypothetical prote	824	6	1.2	333	2	JU0298	Fructose repressor
752	6	1.2	312	2	T43853	conserved hypochet	825	6	1.2	334	2	S00298	Fructose repressor
753	6	1.2	312	2	S50430	dihydrodipicolinat	826	6	1.2	334	2	D85490	Fructose repressor
754	6	1.2	314	2	AD3208	lysophospholipase	827	6	1.2	334	2	D90639	Fructose repressor
755	6	1.2	314	2	T34326	probable peroxidase	828	6	1.2	334	2	AC0517	probable transposo
756	6	1.2	315	2	T09165	hypothetical prote	829	6	1.2	334	2	G85048	serine/chreonine k
757	6	1.2	315	2	A90895	hypothetical prote	830	6	1.2	334	2	A12150	
758	6	1.2	315	2	B96685	hypothetical prote	831	6	1.2	334	2		
759	6	1.2	315	2	B96685	hypothetical prote	832	6	1.2	334	2		

833	6	1.2	335	2	T33483	hypothetical prote
834	6	1.2	335	2	A10481	pekb family carboh
835	6	1.2	336	2	A24430	glyceralddehyde-3-p
836	6	1.2	336	2	G69091	ribosomal protein
837	6	1.2	337	1	A23711	cytochrome-c oxida
838	6	1.2	337	2	S49554	hypothetical prote
839	6	1.2	338	1	C48648	ketol-acid reducto
840	6	1.2	338	2	AG2938	2-hydroxyacid-fam1
841	6	1.2	338	2	H98343	hypothetical prote
842	6	1.2	338	2	G70329	uroporphyrinogen d
843	6	1.2	338	2	C38163	nicotinate-nucleot
844	6	1.2	338	2	H75127	CAAX prenyl protei
845	6	1.2	339	2	AE3389	nicotinate-nucleot
846	6	1.2	339	2	E86393	protein T24P13.2 l
847	6	1.2	340	1	S04898	myb-related protei
848	6	1.2	340	2	E81331	uroporphyrinogen d
849	6	1.2	340	2	C55020	recombination prot
850	6	1.2	340	2	UX0292	recombination prot
851	6	1.2	340	2	E95147	alcohol dehydrogen
852	6	1.2	340	2	A35630	regulatory protein
853	6	1.2	340	2	G70432	conserved hypochet
854	6	1.2	341	1	RPECCT	transcription regu
855	6	1.2	341	2	E91236	transcription regu
856	6	1.2	341	2	E86083	transcription regu
857	6	1.2	341	2	AG0938	transcription regu
858	6	1.2	342	2	T16444	hypothetical prote
859	6	1.2	342	2	T36477	probable DNA-bind
860	6	1.2	342	2	T51703	nicotinate-nucleot
861	6	1.2	342	2	B87701	conserved hypochet
862	6	1.2	343	2	E83673	sorbitol dehydroge
863	6	1.2	343	2	T45415	ketol-acid reducto
864	6	1.2	343	2	D89605	protein F18G5.3 l1
865	6	1.2	343	2	S71094	rad4 protein - Hal
866	6	1.2	343	2	A81423	hypothetical prote
867	6	1.2	344	2	T05987	hypothetical prote
868	6	1.2	344	2	I49585	CD2 antigen protei
869	6	1.2	344	2	A69661	transcription regu
870	6	1.2	344	2	T38745	serine/chreonine p
871	6	1.2	344	2	AD2808	hypothetical prote
872	6	1.2	344	2	S19705	N5,N10-methylene
873	6	1.2	344	2	E95868	probable epoxide h
874	6	1.2	345	1	S58854	cell division prot
875	6	1.2	345	2	AG3186	hypothetical prote
876	6	1.2	345	2	T43736	transcription repr
877	6	1.2	345	2	AC1259	transcription repr
878	6	1.2	345	2	AG1621	transcription repr
879	6	1.2	346	2	F82349	ADP-heptose-Lbs he
880	6	1.2	346	2	T19676	hypothetical prote
881	6	1.2	346	2	AB3057	conserved hypochet
882	6	1.2	346	2	D98229	hypothetical prote
883	6	1.2	346	2	C98015	conserved hypochet
884	6	1.2	347	2	B69019	N5,N10-methylene
885	6	1.2	347	2	T35013	probable membrane
886	6	1.2	347	2	AH2950	conserved hypochet
887	6	1.2	347	2	G96700	protein F12A21.7 l
888	6	1.2	347	2	G84344	hypothetical prote
889	6	1.2	347	2	AB1794	probable rotamase
890	6	1.2	347	2	H64371	malic acid transpo
891	6	1.2	347	2	S40733	hypothetical prote
892	6	1.2	347	2	AF2645	flagellar motor sw
893	6	1.2	347	2	F97427	flagellar motor sw
894	6	1.2	348	2	AG2043	RNA 3'-terminal ph
895	6	1.2	348	2	B81216	peptidyl-prolyl ci
896	6	1.2	348	2	T35968	conserved hypochet
897	6	1.2	348	2	T15219	hypothetical prote
898	6	1.2	348	2	AF3152	hypothetical prote
899	6	1.2	348	2	A12475	hypothetical prote
900	6	1.2	349	2	E75041	hypothetical prote
901	6	1.2	349	2	A71120	hypothetical prote
902	6	1.2	349	2	C72551	probable aminometh
903	6	1.2	350	2	B71473	probable Fe-S clus
904	6	1.2	350	2	S76698	hypothetical prote
905	6	1.2	350	2	T35849	probable ATP/GTP-b
906	6	1.2	350	2	E87327	hypothetical prote
907	6	1.2	351	2	C89788	sorbitol dehydroge
908	6	1.2	351	2	T26918	hypothetical prote
909	6	1.2	351	2	T44428	probable gonococca
910	6	1.2	351	2	AF0975	probable lact fami
911	6	1.2	352	2	T49396	hypothetical prote
912	6	1.2	353	1	FOHWGR	gag polyprotein -
913	6	1.2	353	2	E86604	Fe-S oxidoreductas
914	6	1.2	353	2	F72019	conserved hypochet
915	6	1.2	354	2	D81735	conserved hypochet
916	6	1.2	354	2	E81281	hypothetical prote
917	6	1.2	354	2	AB2844	translation initia
918	6	1.2	354	2	S60967	YGP1 protein precu
919	6	1.2	354	2	T24873	hypothetical prote
920	6	1.2	355	1	BVECMG	UDP-N-acetylglucos
921	6	1.2	355	2	F90640	hypothetical prote
922	6	1.2	355	2	AB1777	UDP-N-acetylglucos
923	6	1.2	355	2	AB1201	UDP-N-acetylglucos
924	6	1.2	355	2	F85491	hypothetical prote
925	6	1.2	355	2	F69298	conserved hypochet
926	6	1.2	355	2	C98332	phc protein (U601
927	6	1.2	356	2	AC0399	2-dehydro-3-deoxy-
928	6	1.2	356	2	A43680	L356 protein - Afr
929	6	1.2	356	2	I39510	carboxylesterase l
930	6	1.2	356	2	T51480	hypothetical prote
931	6	1.2	356	2	T37863	hypothetical sur1-
932	6	1.2	357	2	JO1647	SHL1 protein - hum
933	6	1.2	357	2	B83652	hypothetical prote
934	6	1.2	357	2	T12771	delta-endotoxin ho
935	6	1.2	357	2	AB3371	GTP-binding protei
936	6	1.2	358	1	WMBE38	infected cell prot
937	6	1.2	358	2	T44333	hypothetical prote
938	6	1.2	358	2	B87464	lysophospholipase
939	6	1.2	358	2	AB3249	probable initiatio
940	6	1.2	358	2	H64397	N5,N10-methylene
941	6	1.2	358	2	F69809	spore germination
942	6	1.2	358	2	T00954	hypothetical prote
943	6	1.2	358	2	T48090	hypothetical prote
944	6	1.2	359	2	B40354	arylsulfatylphospha
945	6	1.2	359	2	T41906	hypothetical prote
946	6	1.2	359	2	H69329	nitrate ABC transp
947	6	1.2	359	2	E72290	branched chain am1
948	6	1.2	359	2	SC4738	C02C2.5 protein -
949	6	1.2	359	2	JC7280	cytolec receptor-
950	6	1.2	359	2	A95277	hypothetical prote
951	6	1.2	360	1	G64557	GTP-binding protei
952	6	1.2	360	1	F64601	conserved hypochet
953	6	1.2	360	1	E70816	probable moa2 pro
954	6	1.2	360	2	B71952	hypothetical prote
955	6	1.2	360	2	E71910	hypothetical prote
956	6	1.2	360	2	C72263	motility protein p
957	6	1.2	361	2	JO0374	lignin peroxidase
958	6	1.2	361	2	H70368	conserved hypochet
959	6	1.2	361	2	H70785	probable phosphori
960	6	1.2	361	2	T30402	hypothetical prote
961	6	1.2	361	2	B96978	signal transductio
962	6	1.2	361	2	A96261	hypothetical prote
963	6	1.2	361	2	AB3023	efflux protein lfm
964	6	1.2	362	1	E64793	glycerol dehydroge
965	6	1.2	362	2	F82096	conserved hypochet
966	6	1.2	363	2	T36408	probable esterase
967	6	1.2	363	2	T44150	hypothetical prote
968	6	1.2	364	1	ADHUB	fructose-bisphosph
969	6	1.2	364	2	C84076	hypothetical prote
970	6	1.2	364	2	C70777	probable cobC - My
971	6	1.2	364	2	T47198	H+-exporting ATPas
972	6	1.2	364	2	F90672	probable adhesin l
973	6	1.2	365	2	C69005	succinyl-CoA synth
974	6	1.2	365	2	JC2559	flagellin fljC-1 -
975	6	1.2	365	2	C86746	peptide chain rele
976	6	1.2	365	2	B69114	conserved hypochet
977	6	1.2	365	2	B72458	hypothetical prote
978	6	1.2	365	2	T50566	probable ABC-type

979	6	1.2	365	2	S50409	hypothetical prote
980	6	1.2	366	2	S66016	probable GTP-bind
981	6	1.2	367	2	F84579	probable arginine
982	6	1.2	368	2	C89875	hypothetical prote
983	6	1.2	367	2	T26033	hypothetical prote
984	6	1.2	368	2	C95860	probable alcohol d
985	6	1.2	368	2	T14240	NADH2 dehydrogenas
986	6	1.2	368	2	G70370	conserved hypochet
987	6	1.2	368	2	A41111	flagellum-associat
988	6	1.2	368	2	AB2215	hypothetical prote
989	6	1.2	369	2	E86554	Fe-S oxidoreductas
990	6	1.2	369	2	G72069	conserved hypochet
991	6	1.2	369	2	B71516	probable Fe-S oxid
992	6	1.2	369	2	F81674	conserved hypochet
993	6	1.2	369	2	H82223	nicotinate-nucleot
994	6	1.2	369	2	B85523	hypothetical prote
995	6	1.2	370	1	D22930	DNA repair and gen
996	6	1.2	370	2	D83818	heat-shock protein
997	6	1.2	370	2	G84042	hypothetical prote
998	6	1.2	370	2	H71624	rifin PFB0030c - m
999	6	1.2	370	2	S49583	transcription fact
1000	6	1.2	371	2	D70357	conserved hypochet

ALIGNMENTS

RESULT 1

B81816 nitrite reductase (EC 1.7.99.3) NMA1887 [similarity] - Neisseria meningitidis (strain Z2
C/Species: Neisseria meningitidis
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

R.Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A/Reference number: A81775; MUID:20222556; PMID:10761919

A/Accession: B81816

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-386 <PAR>

A/Cross-references: GB:AL162757; GB:AL157959; NID:97380371; PIDN:CA85110.1; PID:9738052

A/Experimental source: serogroup A, strain Z2491

C/Genetics:

A/Keywords: antiA; NMA1887

C/Keywords: oxidoreductase
F/330,171,179,184/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
F/335,170/Binding site: copper (His) (type 2) (shared with trimetric partner 2) #status P
F/325/Binding site: copper (His) (type 2) (shared with trimetric partner 1) #status predi

Query Match 6.8%; Score 34; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 8e-26; Mismatches 0; Indels 0; Gaps 0;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 YHCAVAPVGMHIANGYGLIVPEKGLPKVDKE 211

DB 169 YHCAVAPVGMHIANGYGLIVPEKGLPKVDKE 202

RESULT 2
E81062 nitrite reductase (EC 1.7.99.3) NMB1623 [similarity] - Neisseria meningitidis (strain MC

N/Alternate names: outer membrane protein Pan1 homolog
C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C/Accession: E81062
R/Retcelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Ok, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Mestiani, V.; Pizze, M.
Science 287, 1809-1815, 2000

A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A/Reference number: A81000; MUID:20175755; PMID:10710307
A/Accession: E81062
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-390 <TET>
A/Cross-references: GB:AE002512; GB:AE002098; NID:97226866; PIDN:AAFA1975.1; PID:972268
A/Experimental source: serogroup B, strain MC58
C/Genetics:
A/Keywords: NMB1623

Query Match 6.8%; Score 34; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 8e-26;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 YHCAVAPVGMHIANGYGLIVPEKGLPKVDKE 211
DB 173 YHCAVAPVGMHIANGYGLIVPEKGLPKVDKE 206

RESULT 3

A49208 nitrite reductase (EC 1.7.99.3) Pan1 [similarity] - Neisseria gonorrhoeae
N/Alternate names: outer membrane protein Pan1
C/Species: Neisseria gonorrhoeae
C/Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 15-Sep-2000

C/Accession: A49208
R/Hoehn, G.T.; Clark, V.L.
Infect. Immun. 60, 4704-4708, 1992

A/Title: The major anaerobically induced outer membrane protein of Neisseria gonorrhoea

A/Reference number: A49208; MUID:93014188; PMID:1198981

A/Accession: A49208

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-392 <HOE>

A/Note: Sequence extracted from NCBI backbone (NCBIP:116468)

C/Keywords: oxidoreductase

Query Match 6.8%; Score 34; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 8.1e-26;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 YHCAVAPVGMHIANGYGLIVPEKGLPKVDKE 211
DB 173 YHCAVAPVGMHIANGYGLIVPEKGLPKVDKE 206

RESULT 4

S32112 (EC 1.7.2.1) (EC 1.7.2.1) - Pseudomonas aureofaciens
C/Species: Pseudomonas aureofaciens
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Jun-2002

C/Accession: S32112
R/Zumft, W.G.
Submitted to the EMBL Data Library, March 1993

A/Reference number: S32112

A/Accession: S32112

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-363 <ZUM>

A/Cross-references: EMBL:Z21945; NID:9287906; PIDN:CAA79939.1; PID:9287907

C/Keywords: copper; oxidoreductase

Query Match 2.4%; Score 12; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPMNVDFHATG 148
DB 111 MPMNVDFHATG 122

RESULT 5

D64744
exopolysaccharide synthetase regulator rcsf - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64744; B47040
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
S.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64740; PMID:97426617; PMID:9278503
A:Accession: D64744
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-134 <BLAT>
A:Cross-references: GB:AE000128; GB:U00096; NID:g1786383; PIDN:AACT3307.1; PID:g1786394;
R:Experimental source: strain K-12, substrain MG1655
R:Gerstein, F.G.; Drapeau, G.R.
J. Bacteriol. 174, 8016-8022, 1992
A:Title: Identification, cloning, and characterization of rcsf, a new regulator gene for
A:Reference number: A47040; PMID:93094132; PMID:1459951
A:Accession: B47040
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-36, DA, 39-45, 47-134 <GER>
A:Cross-references: GB:L04474; NID:g147530; PIDN:AA24508.1; PID:g147532
A:Note: sequence extracted from NCBI backbone (NCBIN:119949, NCBIIP:119951)
C:Genetics:
A:Gene: rcsf
C:Function:
A:Description: stimulates colanic acid capsule synthesis
C:Superfamily: Escherichia coli exopolysaccharide synthetase regulator rcsf

Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 134;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMSGCS 24
DB 10 ALMSGCS 17

RESULT 6
F90653
regulator in colanic acid synthesis Rcsf [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: F90653
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: F90653
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-134 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA33621.1; PID:g13359654; GSPDB:GN00154-
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecsf098
C:Superfamily: Escherichia coli exopolysaccharide synthetase regulator rcsf

Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 134;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMSGCS 24
DB 10 ALMSGCS 17

RESULT 7

F85504
regulator in colanic acid synthesis Rcsf [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
C:Accession: F85504
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grobeck, E.J.; Davis, N.W.; Lin, A.; Dimalanta, E.; Potamoudis, K.; Apodaca
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: F85504
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-134 <STO>
A:Cross-references: GB:AE005174; NID:g12512927; PIDN:AG54498.1; GSPDB:GN00145; UMGCP:20-
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: rcsf
C:Superfamily: Escherichia coli exopolysaccharide synthetase regulator rcsf

Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 134;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMSGCS 24
DB 10 ALMSGCS 17

RESULT 8
J80215
nitrite reductase (EC 1.7.99.3) blue copper-containing - Achromobacter xylosoxidans
N:Alternate names: Nitr
C:Species: Achromobacter xylosoxidans
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 07-May-1999
C:Accession: J80215
R:Vandenberghe, I.H.M.; Meyer, T.E.; Csanovich, M.A.; Van Beeumen, J.J.
Biochem. Biophys. Res. Commun. 247, 734-740, 1998
A:Title: The covalent structure of the blue copper-containing nitrite reductase from Aci
A:Reference number: J80215; PMID:98321197; PMID:9647763
A:Accession: J80215
A:Molecule type: protein
A:Residues: 1-336 <VAN>
A:Comment: This enzyme is responsible for the conversion of nitrite to the gaseous prod
C:Keywords: oxidoreductase

Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 336;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPNVDFF 144
DB 87 MPNVDFF 94

RESULT 9
JG0170
nitrite reductase (EC 1.7.99.3) - Alcaligenes denitrificans subsp. xylosoxydans
C:Species: Alcaligenes denitrificans subsp. xylosoxydans
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0170
R:Suzuki, E.; Horikoshi, N.; Kohzuma, T.
Biochem. Biophys. Res. Commun. 255, 427-431, 1999
A:Title: Cloning, sequencing, and transcriptional studies of the gene encoding copper-c
A:Reference number: JG0170; PMID:99160880; PMID:10049725
A:Accession: JG0170
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <SUZ>
A:Cross-references: GB:AF051831
C:Keywords: copper; oxidoreductase

Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 360;

Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MPMVDFH 144
DB 111 MPMVDFH 118

RESULT 10

D64333
pyruvate synthase (EC 1.2.7.1) alpha chain - Methanococcus jannaschii

C/Species: Methanococcus jannaschii

C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C/Accession: D64333

R/Built: C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Bullock, C.I.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Glodok, A.;

Leon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hartz, M.A.

Science 273, 1058-1073, 1996

A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.

A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A/Reference number: A64300; MUID:96337999; PMID:8688087

A/Accession: D64333

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-389 <BUL>

A/Cross-references: GB:U67482; GB:L77117; NID:g2826267; PIDN:AB98254.1; PID:g1590995; T

C/Genetics:

A/Map position: REV255064-253895

A/Start codon: GTG

C/Superfamily: pyruvate synthase alpha chain; 2-oxoacid ferredoxin oxidoreductase homolo

C/Keywords: coenzyme A; oxidoreductase

Query Match
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 FNKGALGI 347
DB 332 FNKGALGI 339

RESULT 11
F81221
hypothetical protein NMB0240 [imported] - Neisseria meningitidis (strain MCS8 serogroup

C/Species: Neisseria meningitidis

C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C/Accession: F81221

R/Authors: H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.K.; Halt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

rt, H.; Qin, H.; Vamathavan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappapoli, R.; Ve

A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.

A/Reference number: A81000; MUID:20175755; PMID:10710307

A/Accession: F81221

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-483 <TET>

A/Cross-references: GB:AE002381; GB:AE002098; NID:g7225455; PIDN:AA640694.1; PID:g722546

C/Genetics:

A/Experimental source: serogroup B, strain MCS8

C/Superfamily: Neisseria meningitidis probable integral membrane protein NMA0020

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SSTVDAAA 42
DB 440 SSTVDAAA 447

RESULT 12

H81992
probable integral membrane protein NMA0020 [imported] - Neisseria meningitidis (strain

C/Species: Neisseria meningitidis

C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C/Accession: H81992

R/Authors: J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

Nature 404, 502-506, 2000

A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A/Reference number: A81775; MUID:20222556; PMID:10761919

A/Accession: H81992

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-494 <PAR>

A/Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CA83340.1; PID:g73787

A/Experimental source: serogroup A, strain Z2491

C/Genetics:

A/Map position: NMA0020

C/Superfamily: Neisseria meningitidis probable integral membrane protein NMA0020

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SSTVDAAA 42
DB 451 SSTVDAAA 458

RESULT 13
F69802
ABC transporter (ATP-binding protein) homolog yf1c - Bacillus subtilis

C/Species: Bacillus subtilis

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001

C/Accession: F69802

R/Authors: F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Barte

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.T.; Eutian, K.D.; Errington, J.; Fabbri, C.; Ferrari, E.

Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galazzi, A.; Galle

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.

Koetler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogiwara, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon

Rieger, M.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero

A/Authors: Schlach, S.; Schreier, R.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I.

A/Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: F69802

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-604 <KUN>

A/Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CA812651.1; PID:g2633146

A/Experimental source: strain 168

C/Genetics:

A/Gene: yf1c

C/Superfamily: Escherichia coli ABC transporter mlaA; ATP-binding cassette homology

C/Keywords: ATP; nucleotide binding; P-loop

F/382-576/Domain: ATP-binding cassette homology <ABC>

F/399-406/Region: nucleotide-binding motif A (P-loop)

Query Match
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AAATANA 47
DB 474 AAATANA 481

RESULT 14

T48418
ABC transporter-like protein - Arabidopsis thaliana
N/Alternate names: protein P8F6.120
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C/Accession: T48418
R/Author: M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, March 2000
A/Reference number: Z24488
A/Accession: T48418
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-634 <BEV>
A/Cross-references: EMBL:AL162873
A/Experimental source: cultivar Columbia; BAC clone P8F6
C/Genetics:
A/Map position: 5
A/Intons: 197/3; 226/3; 262/3; 351/3; 379/3; 452/1; 480/3; 543/2; 586/3
A/Note: P8F6.120

Query Match

Best Local Similarity 1.6%; Score 8; DB 2; Length 634;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 AAKTANAD 48
|||||

Db 508 AAKTANAD 515

RESULT 15

T21147
hypothetical protein F20D1.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T21147
R/Burton, J.
submitted to the EMBL Data Library, August 1996
A/Reference number: Z19382
A/Accession: T21147
A/Status: preliminary; translated from GB/EMBL/DBD3
A/Molecule type: DNA
A/Residues: 1-915 <WIL>
A/Cross-references: EMBL:Z78542; PIDN:CAB01748.1; GSPDB:GN00028; CESP:F20D1.6
A/Experimental source: clone F20D1
C/Genetics:
A/Gene: CESP:F20D1.6
A/Map position: X
A/Intons: 45/3; 75/3; 126/3; 166/2; 196/3; 243/3; 271/2; 321/1; 388/3; 508/2; 552/1; 70
C/Superfamily: Caenorhabditis elegans hypothetical protein F20D1.6

Query Match

Best Local Similarity 1.6%; Score 8; DB 2; Length 915;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 KMETVEKV 94
|||||

Db 503 KMETVEKV 510

RESULT 16

CCCF55
cytochrome c55 - Chlorobium sp.
C/Species: Chlorobium sp.
C/Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 03-Mar-2000
C/Accession: A00116
R/Van Beeumen, J.; Ambler, R.P.; Meyer, T.E.; Kamen, M.D.; Olson, J.M.; Shaw, E.K.
Biochem. J. 159, 757-774, 1976
A/Title: The amino acid sequences of the cytochromes C-555 from two green sulphur bacteria
A/Reference number: A00116; MUID:77087086; PMID:186412
A/Accession: A00116
A/Molecule type: protein

A/Residues: 1-86 <VAN>

A/Note: the source is designated as Chlorobium thiosulfatophilum
R/Korszun, Z.R.; Saleme, F.R.
Proc. Natl. Acad. Sci. U.S.A. 74, 5244-5247, 1977
A/Title: Structure of cytochrome c55 of Chlorobium thiosulfatophilum: primitive low-po-
A/Reference number: A38043; MUID:78094383; PMID:282997
A/Comment: This basic c-type monoheme cytochrome has been found exclusively in the green
low redox potential compared with mitochondrial cytochrome c. It is reactive with cyto-
C/Superfamily: cytochrome c6; cytochrome c6 homology
C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; photosynthesi
P/4-81/Domain: cytochrome c6 homology <CYC>
P/14,17/Binding site: heme (Cys) (covalent) #status experimental
P/18,60/Binding site: heme iron (His, Met) (axial ligands) #status experimental

Query Match

Best Local Similarity 1.4%; Score 7; DB 1; Length 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 AGKATYD 403
|||||

Db 5 AGKATYD 11

RESULT 17

AG3064
hypothetical protein Atcu4133 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AG3064
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Giller, W.; Grant, C.; Gunthner, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AG3064
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-98 <KUR>
A/Cross-references: GB:AE008689; PIDN:AL44933.1; PID:g17742587; GSPDB:GN00187
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atcu4133
A/Map position: linear chromosome

Query Match

Best Local Similarity 1.4%; Score 7; DB 2; Length 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 RLTVGNG 276
|||||

Db 92 RLTVGNG 98

RESULT 18

S52930
GP41 ENV protein - human immunodeficiency virus type 1 (fragment)
C/Species: human immunodeficiency virus type 1, HIV-1
C/Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 26-Aug-1999
C/Accession: S52930
R/Cohen, J.H.M.; Guetard, D.; Philbert, F.; Chamaret, S.; Taly, T.; Montagnier, L.; d
submitted to the EMBL Data Library, January 1995
A/Description: A novel HIV-1 O-strain illustrates the diversity of the O group.
A/Reference number: S52929
A/Accession: S52930
A/Status: preliminary
A/Molecule type: genomic RNA
A/Residues: 1-104 <COH>
A/Cross-references: EMBL:X84328; NID:g695526; PIDN:CA59066.1; PID:g695527
C/Superfamily: type B retrovirus env polypeptide

Query Match 1.4%; Score 7; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DQDIANY 473
|||||
DB 83 DQDIANY 89

RESULT 19

S16437
ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain - Prochlorothrix hollandica
C/Species: Prochlorothrix hollandica
C/Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C/Accession: S16437

R/Morden, C.W.; Golden, S.S.
J. Mol. Evol. 32, 379-395, 1991
A/Title: Sequence analysis and phylogenetic reconstruction of the genes encoding the large
aroyte Prochlorothrix hollandica.
A/Reference number: S16436; MUID:91251137; PMID:1904095
A/Accession: S16437
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-109 <MOR>

A/Cross-references: EMBL:X57359; NID:945542; PIDN:CAA40633.1; PID:945544
C/Superfamily: ribulose-bisphosphate carboxylase small chain
C/Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 1.4%; Score 7; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 LSDOQIA 471
|||||
DB 19 LSDOQIA 25

RESULT 20

A70407
hypothetical protein aq_1239 - Aquifex aeolicus
C/Species: Aquifex aeolicus
C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jun-2000
C/Accession: A70407

R/Dackert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oy
V.
Nature 392, 353-358, 1998
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A/Reference number: A70300; MUID:98196666; PMID:9537320
A/Accession: A70407

A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-111 <ACF>
A/Cross-references: GB:AE000730; NID:92983674; PIDN:AAC07249.1; PID:92983684; GB:AE00065
A/Experimental source: strain VFS
C/Genetics:

A/Gene: aq_1239
C/Superfamily: Aquifex aeolicus hypothetical protein aq_1239

Query Match 1.4%; Score 7; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 QGLOPFD 234
|||||
DB 28 QGLOPFD 34

RESULT 21

E86906
ribosomal protein S9 [similarity] - Lactococcus lactis subsp. lactis (strain IL1403)
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-May-2002

C/Accession: E86906
R/Bolotin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weisenbach, J.; Ehrlich,
Genome Res. 11, 731-753, 2001
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s

A/Reference number: A86925; MUID:2125186; PMID:11374771
A/Accession: E86906
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-130 <STO>
A/Cross-references: GB:AE005176; PID:g12725323; PIDN:AAK06351.1; GSPDB:GN00146
A/Experimental source: strain IL1403
C/Genetics:

A/Gene: rpsI
C/Superfamily: Escherichia coli ribosomal protein S9

Query Match 1.4%; Score 7; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 GKITYNG 453
|||||
DB 26 GKITYNG 32

RESULT 22

AF3315
transcription elongation factor grea [imported] - Brucella melitensis (strain 16M)
C/Species: Brucella melitensis
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C/Accession: AF3315

R/DalVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
J.; Mazur, M.; Goldman, E.; Selkov, E.; Elser, P.H.; Hagius, S.; O'Callaghan, D.; Lates
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A/Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A/Reference number: AD3252; PMID:11756688
A/Accession: AF3315

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-157 <KUR>

A/Cross-references: GB:AE008917; PIDN:AAL51689.1; PID:g17982422; GSPDB:GN00190
A/Experimental source: strain 16M
C/Genetics:

A/Gene: BMEI0508
A/Map position: I
C/Superfamily: transcription elongation factor greb
C/Keywords: transcription factor

Query Match 1.4%; Score 7; DB 2; Length 157;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 EGDTEV 127
|||||
DB 134 EGDTEV 140

RESULT 23

B97802
hypothetical protein RC0818 [imported] - Rickettsia conorii (strain Malish 7)
C/Species: Rickettsia conorii
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C/Accession: B97802

R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A/Reference number: A97700; MUID:21442074; PMID:11557893
A/Accession: B97802
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-161 <KUR>

A/Cross-references: GB:AE006914; PIDN:AAL03356.1; PID:g15619918; GSPDB:GN00173
C/Genetics:
A/Gene: RC0818

C:Superfamily: nus operon 15K protein

Query Match 1.4%; Score 7; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 QOIANYI 474
|||||
DB 6 QOIANYI 12

RESULT 24

A97742 cytochrome c [imported] - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Nov-2001

C/Accession: A97742

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; Ro

Science 293, 2093-2098, 2001

A>Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A/Reference number: A97700; PMID:21442074; PMID:11557893

A/Accession: A97742

A>Status: preliminary

A/Molecule type: DNA

A/Residues: 1-175 <KUR>

A/Cross-References: GB:AE006914; PIDN:AAL02875.1; PID:G15619399; GSPDB:GN00173

C:Genetics:

A:Gene: CYCM

C:Superfamily: membrane-bound cytochrome c/cym; cytochrome c homology

C/Keywords: chromoprotein; heme; iron; metalloprotein

F:84,87/Binding site: heme (Cys) (covalent) #status predicted

F:88/Binding site: heme iron (His) (axial ligand) #status predicted

F:150/Binding site: heme iron (Met) (axial ligand) #status predicted

Query Match 1.4%; Score 7; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KTNANDN 49
|||||
DB 70 KTNANDN 76

RESULT 25

AH0196 hypoxanthine phosphoribosyl transferase [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C/Accession: AH0196

R:Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;

Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; PMID:21470413; PMID:11586360

A/Accession: AH0196

A>Status: preliminary

A/Molecule type: DNA

A/Residues: 1-191 <KUR>

A/Cross-References: GB:AL590842; PIDN:CAC90435.1; PID:G15979651; GSPDB:GN00175

C:Genetics:

A:Gene: YPO1613

Query Match 1.4%; Score 7; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 EYPPVPD 77
|||||
DB 40 EYPPVPD 46

RESULT 26

T29645 hypothetical protein W01A11.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000

C/Accession: T29645

R:Blanchard, M.; Bradshaw, H.

submitted to the EMBL Data Library, July 1996

A/Description: The sequence of C. elegans cosmid W01A11.

A/Reference number: Z20658

A/Accession: T29645

A>Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-192 <BLA>

A/Cross-References: EMBL:U64852; PIDN:AB04967.1; GSPDB:GN00023; CESP:W01A11.4

A/Experimental source: strain Bristol N2; clone W01A11

C:Genetics:

A:Gene: W01A11.4

A/Map position: 5

A/Intons: 7/3, 101/3

C:Superfamily: beta-galactoside-binding lectin

Query Match 1.4%; Score 7; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 DRVHEEG 298
|||||
DB 133 DRVHEEG 139

RESULT 27

F64941 rnd protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002

C/Accession: F64941; S41588

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; PMID:97426617; PMID:9278503

A/Accession: F64941

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-193 <BLAT>

A/Cross-References: GB:AE000275; GB:U00096; NID:G1788106; PIDN:ACT4876.1; PID:G1788108

A/Experimental source: strain K-12, substrain MG1655

R:Paula, M.; Heinz, E.; Wolter, F.P.

Mol. Gen. Genet. 242, 241-249, 1994

A>Title: The fadD gene of Escherichia coli K12 is located close to rnd at 39.6 min of t

A/Reference number: S41588; PMID:94150456; PMID:8107670

A/Accession: S41588

A/Molecule type: DNA

A/Residues: 121-123, 'P', 125-193 <FUD>

C:Genetics:

A:Gene: rnd

C:Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23
|||||
DB 17 ALMLSGC 23

RESULT 28

C90943

probable outer membrane protein ECG2515 [imported] - Escherichia coli (strain O157:H7,

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C/Accession: C90943

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
 A/Reference number: A96629; MUID:21156233; PMID:11258796
 A/Accession: C90943
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-193 <HAV>
 A/Cross-references: GB:BA000007; PIDN:BA35938.1; PID:G13361982; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R1MD 0509952
 C/Genetics:
 A/Gene: EC82515
 C/Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSCG 23
 |||||
 Db 17 ALMLSCG 23

RESULT 29
 G85791
 Probable outer membrane protein Z2849 [imported] - *Escherichia coli* (strain O157:H7, sub
 C/Species: *Escherichia coli*
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: G85791
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobocck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: G85791
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-193 <STO>
 A/Cross-references: GB:AB005174; NID:G12515846; PIDN:AAG56795.1; GSPDB:GN00145; UMGP:Z28
 C/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: Z2849
 C/Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSCG 23
 |||||
 Db 17 ALMLSCG 23

RESULT 30
 AD0725
 Probable lipoprotein STY1949 [imported] - *Salmonella enterica* subsp. *enterica* serovar Ty
 C/Species: *Salmonella enterica* subsp. *enterica* serovar Typh
 A/Note: this species has also been called *Salmonella typh*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: AD0725
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
 th, T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A/Authors: Parky, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A>Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
 A/Reference number: AB0502; MUID:21534947; PMID:11677608
 A/Accession: AD0725
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-193 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD05502.1; PID:G16503006; GSPDB:GN00176

C/Genetics:
 A/Gene: STY1949
 C/Superfamily: rnd protein

Query Match 1.4%; Score 7; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSCG 23
 |||||
 Db 17 ALMLSCG 23

RESULT 31
 G69539
 ribosomal protein L15, cytosolic - *Archaeoglobus fulgidus*
 C/Species: *Archaeoglobus fulgidus*
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
 C/Accession: G69539
 R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A/Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Ariach, P.; Kaime, B.P.; Sykes, S
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
 A/Reference number: A69250; MUID:98049343; PMID:9389475
 A/Accession: G69539
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-194 <LEB>
 A/Cross-references: GB:AE00944; GB:AE000782; NID:G2689267; PIDN:AAB88937.1; PID:G26482
 C/Superfamily: rat ribosomal protein L15
 C/Keywords: cytosol; nucleotide binding; P-loop; protein biosynthesis; ribosome
 P160-168/Region: nucleotide-binding motif A (P-loop)

Query Match 1.4%; Score 7; DB 2; Length 194;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 VDRDHPA 82
 |||||
 Db 134 VDRDHPA 140

RESULT 32
 T36138
 hypothetical protein SCE19A.15c - *Streptomyces coelicolor*
 C/Species: *Streptomyces coelicolor*
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000
 C/Accession: T36138
 R;Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A/Reference number: Z21598
 A/Accession: T36138
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-202 <SEB>
 A/Cross-references: EMBL:AL096857; PIDN:CAB50996.1; GSPDB:GN000070; SCOEDB:SCE19A.15c
 A/Experimental source: strain A3(2)
 C/Genetics:
 A/Gene: SCOEDB:SCE19A.15c
 C/Superfamily: *Streptomyces coelicolor* hypothetical protein SCE19A.15c

Query Match 1.4%; Score 7; DB 2; Length 202;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 PEGAPOA 376
 |||||
 Db 77 PEGAPOA 83

RESULT 33
A:Accession: A85098
A:Title: hypothetical protein AT4g09590 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: A85098
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: A85098
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <STO>
A:Cross-references: GB:NC_001268; NID:g7267654; PIDN:CA878082.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g09590
A:Map position: 4

Query Match
Best Local Similarity 1.4%; Score 7; DB 2; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 NQADKAA 31
DB 184 NQADKAA 190

RESULT 34
C84775
A:Accession: C84775
A:Title: Probable harpin-induced protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Molnar, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
Eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84775
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <STO>
A:Cross-references: GB:AE002093; NID:g4510372; PIDN:AAD21460.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g35970
A:Map position: 2

Query Match
Best Local Similarity 1.4%; Score 7; DB 2; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 NQADKAA 31
DB 184 NQADKAA 190

RESULT 35
C70737
A:Accession: C70737
A:Title: Probable sigd protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Conroy, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Bartell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70737
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-212 <COL>
A:Cross-references: GB:Z77165; GB:AL123456; NID:g7261609; PIDN:CA801009.1; PID:e255083;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: sigd

Query Match
Best Local Similarity 1.4%; Score 7; DB 2; Length 212;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 488 LSADPVA 494
DB 71 LSADPVA 77

RESULT 36
T29906
A:Accession: T29906
A:Title: hypothetical protein F59A3.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29906
R:Wu, X.; Le, T.T.
Submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid F59A3.
A:Reference number: Z20707
A:Accession: T29906
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-212 <WUX>
A:Cross-references: EMBL:U80448; PIDN:AA837817.1; GSPDB:GN000019; CESP:F59A3.5
A:Experimental source: strain Bristol N2; clone F59A3
C:Genetics:
A:Gene: CESP:F59A3.5
A:Map position: 1
A:Insertions: 12/1; 43/3; 78/3; 108/3; 155/1

Query Match
Best Local Similarity 1.4%; Score 7; DB 2; Length 212;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 ALKATVG 266
DB 104 ALKATVG 110

RESULT 37
S55925
A:Accession: S55925
A:Title: Probable arabinogalactan protein precursor - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S55925
R:Fogson, B.O.; Davies, C.
Plant Mol. Biol. 28, 347-352, 1995
A:Title: Characterization of a cDNA encoding the protein moiety of a putative arabinogal
A:Reference number: S55925; MUID:95322597; PMID:7599320
A:Accession: S55925
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-215 <POG>
A:Cross-references: EMBL:247980; NID:g872126; PIDN:CAA8023.1; PID:g872127
C:Superfamily: proline-rich protein

Query Match
Best Local Similarity 1.4%; Score 7; DB 2; Length 215;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 PKTPAPA 389
DB 45 PKTPAPA 51

RESULT 38

G91207
 Probable replicase [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002
 C:Accession: G91207
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasaewara, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Rep. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A:Reference number: A96629; MUID:2156231; PMID:11258796
 A:Accession: G91207
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-218 <HAV>
 A:Cross-references: GB:BA000007; PIDN:BA838054.1; PID:G13364106; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Gene: Ec84631
 C:Superfamily: Escherichia coli hypothetical protein y1dx

Query Match 1.4%; Score 7; DB 2; Length 218;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALMLSGC 23
 |||||
 Db 18 ALMLSGC 24

RESULT 39

T15381
 hypothetical protein C03B1.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T15381
 R:Martin, J.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C03B1.
 A:Reference number: Z18340
 A:Accession: T15381
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-230 <MAR>
 A:Cross-references: EMBL:U40952; NID:G1072237; PID:G1072239; PIDN:AAA81738.1; CESP:C03B1
 C:Genetics:
 A:Gene::CESP:C03B1.9
 A:Introns: 29/1; 116/1; 134/3; 192/3

Query Match 1.4%; Score 7; DB 2; Length 230;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 ITVGNQ 455
 |||||
 Db 58 ITVGNQ 64

RESULT 40

I38849
 LERK-3 - human
 C:Species: Homo sapiens (man)
 C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 29-Sep-1999
 C:Accession: I38849
 R:Kozlosky, C.J.; Maraskovsky, E.; McGrew, J.T.; VandenBos, T.; Teepe, M.; Lyman, S.D.;
 Oncogene 10, 299-306, 1995
 A>Title: Ligands for the receptor tyrosine kinases hek and elk: isolation of cDNAs encod
 A:Reference number: I38849; MUID:5140419; PMID:7838529
 A:Accession: I38849
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-238 <RES>
 A:Cross-references: EMBL:U14187; NID:G642832; PIDN:AAC50078.1; PID:G642833
 C:Genetics:

A:Gene: GDB:EPLG3
 A:Cross-references: GDB:438336; OMIM:601381
 A:Map position: 1q21-1q22
 C:Superfamily: axon guidance signal protein

Query Match 1.4%; Score 7; DB 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
 |||||
 Db 78 GPGGAE 84

Search completed: August 27, 2003, 18:45:53
 Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 5, 2003, 08:38:48 / Search time 5500 Seconds
(without alignments)
3733.933 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626
Sequence: 1 MSKPTLITKTTTICALSALML.....NKGQLSADVDVAKAKTKRPN 502

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045461386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPRO/spool/US10088045/runat_04092003_083142_3916/app_query.fasta_1.647
-DB=genemb1 -QFMT=fastcap -SUFFIX=rg -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALLIG=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10088045 @CGCN 1.1 3508 -runat_04092003_083142_3916 -NCPU=6 -ICPU=3
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_pl:*
25: em_pb:*
26: em_ro:*
27: em_scs:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2626	100.0	1506	6 AX098868	AX098868 Sequence
2	2626	100.0	1509	6 AX098866	AX098866 Sequence
3	2626	100.0	1759	6 AX394864	AX394864 Sequence
4	2626	100.0	66986	6 AX067454	AX067454 Sequence
5	1495.5	56.9	166050	1 AL646085	AL646085 Ralstonia
6	1394	53.1	9792	1 AE002512	AE002512 Neisseria
7	1394	53.1	326501	1 NMA622491	AL162757 Neisseria
8	1394	53.1	349980	6 NC044033	AX044033 Sequence
9	1392	53.0	2069	1 NC00PANI	M97926 Neisseria g
10	787	30.0	3675	1 AB076606	AB076606 Halorubrum
11	759	28.9	2506	1 HMA278286	AJ278286 Halorubrum
12	754.5	28.7	1439	1 HD557012	AJ557012 Halorubrum
13	494.5	18.8	1823	1 AB046603	AB046603 Alcaligen
14	486	18.5	2375	1 AB013078	AB013078 Alcaligen
15	474	18.1	1307	1 RSU62291	U62291 Rhodococcus
16	469.5	17.9	1903	1 ACNITRD	Z48635 A. cycloclac
17	456	17.4	10587	1 AE007256	AE007256 Sinorhiz
18	453	17.3	1886	1 PSENIRAX	M97994 Pseudomonas
19	453	17.3	4029	1 AF040987	AF040987 Pseudomon
20	453	17.3	10484	1 AE008247	AE008247 Agrobacte
21	453	17.3	10753	1 AE009367	AE009367 Agrobacte
22	453	17.3	20761	1 AF083948	AF083948 Pseudomon
23	452	17.2	1709	1 AF051831	AF051831 Alcaligen
24	451	17.2	1143	1 PANIRKA	Z21945 P. aureofaci
25	442	16.8	298550	1 AP005961	AP005961 Bradyrhiz
26	440	16.8	1499	1 RHU65658	U65658 Rhizobium h
27	436.5	16.6	2147	1 BUJ2516	AJ002516 Bradyrhiz
28	431	16.4	11630	1 AE009732	AE009732 Brucella
29	429	16.3	11302	1 AE014526	AE014526 Brucella
30	428.5	16.3	316050	1 BX321859	BX321859 Nitrosomo
31	427.5	16.3	2351	1 AEPANIR	D13155 A. faecalis
32	390.5	14.9	313200	1 AP005214	AP005214 Corynebac
33	258	9.8	10159	1 AE005046	AE005046 Halobacte
34	258	9.8	10485	1 AE011334	AE011334 Leptospir
35	249	9.5	537	1 AF339049	AF339049 Nitrosomo
36	248.5	9.5	343250	1 AP003594	AP003594 Nostoc sp
37	245	9.3	347456	1 AE017003	AE017003 Bacillus
38	232	8.8	537	1 AY078254	AY078254 Mesorhizo
39	229	8.7	537	1 AY078247	AY078247 Ensifer s
40	228	8.7	537	1 AY078248	AY078248 Ensifer s
41	227	8.6	299600	1 AP005942	AP005942 Bradyrhiz
42	226.5	8.6	540	1 AF339044	AF339044 Nitrosomo
43	224.5	8.5	540	1 AF339046	AF339046 Nitrosomo
44	224.5	8.5	540	1 AF339047	AF339047 Nitrosomo
45	224	8.5	537	1 AY078255	AY078255 Pseudomon

RESULT 1

ALIGNMENTS

AX098868
LOCUS AX098868 1506 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 3 from Patent WO0119996.
ACCESSION AX098868
VERSION AX098868.1 GI:13538108
KEYWORDS
SOURCE Moraxella catarrhalis
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.
REFERENCE
1
AUTHORS Thomard, J.
TITLE Moraxella catarrhalis antigen, corresponding gene and uses thereof
JOURNAL Patent: WO 0119996-A 3 22-MAR-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
source location/Qualifiers
1.1506
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"
BASE COUNT 431 a 356 c 363 g 356 t
ORIGIN
Alignment Scores:
Pred. No.: 1,18e-166 Length: 1506
Score: 2626.00 Matches: 502
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-088-045-2 (1-502) x AX098868 (1-1506)
QY 1 MetSerTyrProThrLeuIleuValThrThrLeuIleuValSerAlaLeuMetLeu 20
DB 1 ATGTCTAAGCTTACTTGTGTAATAAACAACCTTAATTTGCTTAAGTGCATTGATGCTC 60
QY 21 SerGlyCyseSerAnGlnAlaAspLysValAlaGlnProLysSerThrValAspAla 40
DB 61 AGTGGTTGTAGCATCAATCAAGCGAACAAGCCGCCAGCAAAAAGCAGCAGTGAAGCT 120
QY 41 AlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnIleGlnGlyLeuPro 60
DB 121 GCCGCCAAGACACCAATGATGATTAATGCTGCTTCAAGAACATCAAGCGAGCTGCT 180
QY 61 ValIleAspAlaIleValThrHisAlaProGlnValProProValAspAsnHis 80
DB 181 GTATTGATGTCATTTGATGACATGACACGAAGTTCCACACCTGTTGACCGTGAACAC 240
QY 81 ProAlaLysValValIleLysMetGluThrValGlnLysValMetArgLeuAlaAspGly 100
DB 241 CCGGCCAAAGTGGTGGTAAAAATGGAACCGTTGAAATAATGCTGCTGCGACATGCG 300
QY 101 ValGlnTyrGlnPheThrPheGlnGlyGlnValProGlnMetIleArgValArg 120
DB 301 GTGGAAATATCAGTTTGGATTTGGCGGTCAAGTTCCAGGCGAGATGATTCGTGGCG 360
QY 121 GlnGlyAspThrIleGlnValGlnPheSerAsnHisProAspSerLysMetProHis 140
DB 361 GAAAGGGAACCAATCGAAGTGAAGTTCTCAACACCCAGATTCAAAAATGCCCATAA 420
QY 141 ValAspPheHisAlaAlaThrGlyProGlyGlyValAlaGlnLysSerPheThrAlaPro 160
DB 421 GTTGAATTTACAGCTGCCACAGGCGCTGGCGGCGGAGAACGCTCATTTACCGACCG 480
QY 161 GlyHisThrSerThrPheSerPheLysValAlaGlnProGlyLeuTyrValTyrHis 180
DB 481 GGTTCATACATCAACCTTTAGTTTAAGCCTTACAGCTGTTGATGCTATCACTGA 540
QY 181 AlaValAlaProValGlnMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGln 200
DB 541 GGGGTTGGCCCTGTTGGCATGACATTTGTAATGGCATGATGATTTGATTTGGTTGAA 600

QY 201 ProLysGlnGlyLeuProLysValAspLysGlnTyrTyrValMetGlnGlyAspPheTyr 220
DB 601 CCAAAAAGAGGCGCTTGCCTCAAAAGATGAATAAAGATCTATGCTATGCAAGCGCATTTAT 660
QY 221 ThrLysGlyLysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGlnLysAlaIleArg 240
DB 661 ACCAAAGGCAAAATATGCGCAACAGGCTTACAGCCCTTGTATGATAAAGCCATTCGA 720
QY 241 GlnAspAlaGlnTyrValValPheAsnGlySerValAlaLeuThrGlyGlnAsnAla 260
DB 721 GAAAGATGCTGAATATGTTGCTTAATGATGCGTGGGCGCTTGAAGTGAATAATGCT 780
QY 261 LeuLysAlaLysValGlyGlnThrValAlaGlnPheValAlaGlnGlyProAsnLeu 280
DB 781 CTAAAGCCCAAGTTGGCAAACTGTTCCCTTATTTGTGGTACCGCGGCCGAATTTG 840
QY 281 ThrSerSerPheHisValIleGlyGlnIlePheAspLysValHisPheGlnGlyGlyLys 300
DB 841 ACATCATCATTCATCATGCTATGCTGATGATTTGATTAAGTTCACTTGAAGGCTGAAG 900
QY 301 GlyLysAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValAlaIleThrGln 320
DB 901 GGTGAACCAACCAATATCCAAACCAACGCTAATCCAGAGTGGCGCTCCATCACTGA 960
QY 321 PheLysValAspValProGlyAspTyrValIleValAspHisAlaIlePheArgAlaPhe 340
DB 961 TTTAAGTGTGATGTGCGGGGTATTAATGCTTGTGTTGACCAAGCCATCTTCGCTGCT 1020
QY 341 AsnLysGlyValAlaLeuGlyTyrLeuLysValGlnGlyGlnAsnHisGlnIleTyrSer 360
DB 1021 AACAAAGGGGCAATGGGCACTTAAGTGAAGGGAAGAAATCATGATGATTTATCA 1080
QY 361 HisLysGlnThrAspAlaValTyrLeuProGlnGlyValProGlnAlaIleAspThrGln 380
DB 1081 CACAAACCAACAGACGCTGTATCTGACAGGGGCGCCCAACCAATGATATACCA 1140
QY 381 GlnAlaProLysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAla 400
DB 1141 GAAAGCAACCAAAACCTGACCTGCACTTCAAGACAGATTAACAGATTAAGCA 1200
QY 401 ThrTyrAspSerAsnGlyAlaLysHisGlnProAspGlyLysGlyValProAsnAla 420
DB 1201 ACCATGATCTTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaIleSerIle 440
DB 1261 TTCCACAGCGCTTGCAACTGTATCTGAAGCGCCGACAGCTGCTGCGCAGCATC 1320
QY 441 ValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlnLysSerVal 460
DB 1321 GTGGCAATGATTTGCTGTGAAGATTACCGTCAATGGCAACCAATATGAAGCGTCA 1380
QY 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuAsnSer 480
DB 1381 CCTGGGATTTGCTGTAGCAACCAACAGATTTGCCAATGTCATACCTACACGCTTAAC 1440
QY 481 PheGlyAsnLysGlyGlnLysSerAlaAspAspValAlaLysAlaLysLysThrLys 500
DB 1441 TTTGGTAAACAAAGGCGTCACTCACTGACGATGTCGCAAAAGCCAAAACCAAG 1500
QY 501 ProAsn 502
DB 1501 CCAAAC 1506
RESULT 2
AX098868 1509 bp DNA linear PAT 02-APR-2001
LOCUS AX098868
DEFINITION Sequence 1 from Patent WO0119996.
ACCESSION AX098868
VERSION AX098868.1 GI:13538107
KEYWORDS Moraxella catarrhalis
SOURCE Moraxella catarrhalis
ORGANISM Moraxella catarrhalis

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.

REFERENCE
1 Thonard, J.

AUTHORS
TITLE
JOURNAL
Patent: WO 0119996-A 1 22-MAR-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)

FEATURES
source
1. 1509
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"

BASE COUNT 432 a. 356 c 364 g 357 t
ORIGIN

Alignment Scores:

Score: 1,186-166 Length: 1509
Percent Similarity: 2626.00 Matches: 502
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: 100.00% Indels: 0
Gaps: 0

US-10-088-045-2 (1-502) x AX098866 (1-1509)

QY 1 MetSerLySPProThrLeuLleuLyThrThrLeuLleCyAlaLeuSerAlaLeuMetLeu 20
DB 1 AAGTCAAGACCTACTTGTATTAACACCTTAATTTGTGCTTAAGTCACTTGAATGC 60
QY 21 SerGlyCySerSerAsnGlnAlaAspLyAlaAlaGlnProLySerSerThrValaAla 40
DB 61 AGTGGTTGTAGCAATCAACCGGCAAGCGCCACCAAAAAGCAGCGTAGACGCT 120
QY 41 AlaAlaLyThrAlaAsnAlaAspAlaAlaSerGlnGlnGlnGlnGlnGlnGlnGln 60
DB 121 GCCGCCAAGACAGCAATGCAATGATGCTCTCAAGAACATCAAGCGACGCTGCT 180
QY 61 ValAlaSerAlaAlaLeuAlaThrHisAlaProGluValaProProValaAspArgAla 80
DB 181 GTCAATTGATGCCATTGTTGCGATGCAACCAAGTTCCACCTGTTGACCGTGACAC 240
QY 81 ProAlaLyValaValaValaLySerGluThrValaGlnLyValaMetArgLeuAlaAspGly 100
DB 241 CCCGCCAAAGTGTGGTAAATAATGGAACCGTTGAAATAATGCTGCTGCGCAGATGC 300
QY 101 ValGluTyrGlnPheThrPheGlyGlnValaProGlyGlnMetIleArgValaArg 120
DB 301 GTGGAATATCAGTTTGGACATTTGGCGGTCAAGGTCAGGCGCATGATGCTGTGCGT 360
QY 121 GlnGlyAspThrIleGluValaGlnPheSerAsnHisProAspSerLyMetProHisAsn 140
DB 361 GAAGCGACACCATGCAAGTGCAGTTCTCAACCAACCAAGTTCAAAAATGCCCATAT 420
QY 141 ValaAspPheHisAlaAlaThrGlyProGlyGlyAlaGlnAlaSerPheThrAlaPro 160
DB 421 GTTGACTTTCAAGCTCCACAGGGCTGGGGGGGAGAGAGCGTCATTATACCGACCG 480
QY 161 GlyHisThrSerThrPheSerPheLyAlaLeuGlnProGlyLeuTyrValaTyrHisGly 180
DB 481 GGTCAATACCAACTTTAGTTTAAAGCTTACACCTGTTTGTATGCTATACACTGT 540
QY 181 AlaValaAlaProValaGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValaGlu 200
DB 541 GCGGTGGCCCTGTTGGCATGCACTGCTTAAGCATGATAGTGTGATTTGGTTGA 600
QY 201 ProLySerGlnGlyLeuProLyValaAspLyGlyTyrTyrValaMetGlnGlyAspPheTyr 220
DB 601 CCAAAAGAGGGCTTGCCTAAAGATGATTAAGATCTATGTCAGCAAGCGCACTTTAT 660
QY 221 ThrLySerGlyTyrTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 240
DB 661 ACCAAAGGCAATATGCGCAAGGCTTACAGCCCTTGTATATGGAATAAAGCACTTCA 720

QY 241 GluAspAlaGluTyrValaValaPheAsnGlySerValaGlyAlaLeuThrGlyGlyAla 260
DB 721 GAAGATGCTGAATATGTTGTCTTTAATGTTTCGGGGGGCGTTGACTGTGCAAAATGCT 780
QY 261 LeuLyAlaLyValaGlyGluThrValaArgLeuPheValaGlyAsnGlyGlyProAsnLeu 280
DB 781 CTAAGAGCCAGGTGGGGAACCTGTGCTTATTTGGGTTAAGCGGCGCCGAATTTTG 840
QY 281 ThrSerSerPheHisValaIleGlyGlnIlePheAspLyValaHisPheGlyGlyGly 300
DB 841 ACATCATCATTCATGCTGATGATGATTTTGTATTAAGTTCACTTTGAGGGTGTAAAG 900
QY 301 GlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValaAlaIleThrGlu 320
DB 901 GGTGAAAACCAATATCCAAACCAAGCTTAATCCAGAGGTGGCGCTGCCATCTGAA 960
QY 321 PheLyValaAspValaProGlyAspTyrValaLeuValaAspHisAlaIlePheArgAlaPhe 340
DB 961 TTTAAGGTGATGTCGGGTGATATGCTTGGTTGACCATGCACTTCCGTCCTTT 1020
QY 341 AsnLySerGlyAlaLeuGlyIleLeuLyValaGlnGlyGlnGlnGlnGlnGlnGlnGln 360
DB 1021 AACCAAGGGGCACTTGGCATACTTAAGTGGAGGTGAAGAAATCATGAGATTTATCA 1080
QY 361 HisLySerGlnThrAspAlaValaTyrLeuProGluGlyAlaProGlnAlaIleAspThrGln 380
DB 1081 CACAACAACAACGACGCTGTATCTGCGACAGGGTGGCCCAAGCAATGATACCCAA 1140
QY 381 GluAlaProLySerThrProAlaProAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 400
DB 1141 GAAGACCCCAAAACCTGCACTGCACTTCAACAGACCAATTAAGACAGATTAAGCA 1200
QY 401 ThrTyrAspSerAsnCyAlaAlaIleCysHisGlnProAspGlyLySerGlyValaProAsnAla 420
DB 1201 ACTTATGACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIle 440
DB 1261 TTCACACCGCTTGCACACTGATCTGTAACGCGGCAACGCGCTGTCGCCCGACATC 1320
QY 441 ValaAlaAsnGlyLeuSerGlyTyrValaIleThrValaAsnGlyAsnGlyGlnSerValaMet 460
DB 1321 GTGGCAATATGATTTCTGTTAAGATTACCGCAATGCGCAATATGAAGCCGTCAG 1380
QY 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValaIleThrTyrThrLeuAsnSer 480
DB 1381 CCTGCGATTTGCTTGAAGCAACCAAGATGCGCAATGCTATCACTTACCGCTTAACGC 1440
QY 481 PheGlyAsnLySerGlyGlyGlnLeuSerAlaAspAspValaAlaLyAlaLySerThrLy 500
DB 1441 TTTGGTAAACAAAGCGCTCAACTCAGTGCAGACGATGTGGCAAAAGCCAAAACCAAG 1500
QY 501 ProAsn 502
DB 1501 CCAAAC 1506
RESULT 3
AX394864 1759 bp DNA linear PAT 18-MAY-2002
LOCUS AX394864
DEFINITION Sequence 1 from Patent WO0218595.
ACCESSION AX394864
VERSION AX394864.1 GI:21065937
KEYWORDS
SOURCE
ORGANISM
Moraxella catarrhalis
Moraxella catarrhalis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.
REFERENCE
1 Iosmore, S., Wang, J., Bradley, B., Ochs, M., and Yang, Y. P.
AUTHORS
TITLE
JOURNAL
Moraxella polypeptides and corresponding dna fragments and uses thereof
Patent: WO 0218595-A 1 07-MAR-2002;
Aventis Pasteur Limited (CA)

FEATURES

Source

CDS

Location/Qualifiers

1..1759

/organism="Moraxella catarrhalis"

/mol_type="genomic DNA"

/db_xref="taxon:480"

101..1609

/note="unnamed protein product"

/codon_start=1

/transl_table=1

/protein_id="CAD32058.1"

/db_xref="GI:21065938"

translation="MSKPTIKTTLICALSALMLSGSNQADKAPKSTVDAAKT
 AANDNASQEHQBLEVIDAVTHAPVPPVDHDPKAVVMEVEKWRADVE
 YQFTGQVPGQIMIRREGDTI EVOFSNPDSPKPNVDFAHATPGGGAESFTAP
 GHTSTSEFKALQGLVYVHCVA PVGWHINLNGMYGLIVPEKGLPKVDKEYVMGD
 FYTKKGEQGLQFPDMKEKALREDAEYVNGSVGLATGENALKAKVGETVRLVNGG
 GPLTSEFHVIGELPDKVHEGKGEMHNIOTLLIPAGSAITLPEFVVDGVDVLDH
 AIRAPRNKALGILKVEENHETYSKQTDAYLPEGAQALIDTQAPTPAPALQ
 EQIKAGATYDSNCAOHOPDGKVPAPFPPLANSIDYLNADHARLAASIVANGLSKIT
 VNGQYESVMPALALSDQAIANVITYTLNLSFGNKGQLSADVAKAKTKPN"

BASE COUNT 497 a 396 c 414 g 452 t
 ORIGIN

Alignment Scores:

Pred. No.: 1,41e-166 Length: 1759
 Score: 2626.00 Matches: 502
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-088-045-2 (1-502) x AX394864 (1-1759)

QY 1 MetSerLyProThrLeuIleLyThrThreIleCyAlaLeuSerAlaLeuMetLeu 20
 DB 101 ATGCTTAAGCTTACTTTGATTAACCACTTAATTTTGCTTAACTGATGATGCTC 160
 QY 21 SerGlyCySerSerAnginaIaAspLyAlaAlaGlnProLySerSerThrValaAspAla 40
 DB 161 AGTGGTGTGATGATCAATCAAGCGGCAAAAGCCGCCAGCAAAAAGCAGCATGAGAGCT 220
 QY 41 AlaAlaLyThrAlaAlaAlaAspAlaAlaSerGlnGlnIleGlnGlyLeuPro 60
 DB 221 GCGGCCAAGACACCAATGAGATTAATGCTGCTCAACAAACATCAAGCGGAGCTGCC 280
 QY 61 ValIleAspAlaIleValThrAlaAlaProGluValProProProValaAspAspHis 80
 DB 281 GTCTATGATGCTGATGTCATGACGACCAAGAGTTCCACCACTGTTGACCGTGACAC 340
 QY 81 ProAlaLyValaValaLySerGluThrValaGluLyValaMetArgLeuAlaAspGly 100
 DB 341 CCGGCCAAGGTGGGTGTAATAATGAAACCGTTGAATAAGTCAATGCTGCGCAGATGGC 400
 QY 101 ValGluLyGlnPheThrPheThrPheGlyGlyGlnValaProGlyGlnMetIleArgValaArg 120
 DB 401 GTGGAATATAGTTTGACATTTGGCGGTCAAGTTCCAGGCGAGATGATTCGTGTGCGGT 460
 QY 121 GluGlyAspThrIleGluValaGlnPheSerAsnHisProAspSerLySerMetProHisAsn 140
 DB 461 GAAGGCGACACCATCGAAGTGCAGTTCTCAAAACCAACCAATTCAAAATGCCCCATTAAT 520
 QY 141 ValaAspPheHisAlaAlaThrGlyProGlyGlyValaGluAlaSerPheThrAlaPro 160
 DB 521 GTTGACTTTCACCGTGCACAGGGCGTGGCGGGCGGACAAACGTCATTTACCGCACCG 580
 QY 161 GlyHisThrSerThrPheSerPheLyAlaLeuGlnProGlyLeuTyValaIleHisGlyArg 180
 DB 581 GGTCAATACATCAACCTTTAGTTTAAAGCCTTACAGCTGTGTTGATGCTATCACTGT 640
 QY 181 AlaValaAlaProValaGlyMetHisAlaAlaAsnGlyMetTyThrGlyLeuIleLeuValaGlu 200
 DB 641 GCGGTGGCCCTGTGTGCATGACATTTGTAATGGCATGTATGATTTGTTGTTGA 700

QY 201 ProLyGluGlyLeuProLyValaAspLyGluTyTyTyValaMetGlnIleAspPheTyR 220
 DB 701 CCAAAAAGAGCGCTTGCACAAAGTAAAGATTAATGATGTCATGCAAGCGCATTTTAT 760
 QY 221 ThrLyGlyLyTyTyGlyGluGlnIlePheAspMetGlnLyValaIleArg 240
 DB 761 ACCAAAAGGCAATATGCGCAACAGGCTACAGCCCTTTGATATGGAATAAGCCATTGCA 820
 QY 241 GluAspAlaGluTyTyValaValaPheAsnGlySerValaGlyAlaLeuThrGlyGluAsnAla 260
 DB 821 GAAGATGCTGATATGTTGTTCTTTAAATGTTGCGTGGGGCGTTGATGTAATAATGCT 880
 QY 261 LeuLyAlaLyValaGlyGluThrValaArgLeuPheValaGlyAsnGlyGlyProAsnLeu 280
 DB 881 CTAAAAAGCCAAAGTTGGCCAAACTGTTCCCTTATTTGTGGTAAAGCGCGCCGAATTGG 940
 QY 281 ThrSerSerPheHisValaIleGlyGluIlePheAspLyValaHisPheGluGlyGlyArg 300
 DB 941 ACATCATCTTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 1000
 QY 301 GlyGluAsnHisAsnIleGlnThrThreIleProAlaGlyGlyAlaIleThrGlu 320
 DB 1001 GGTGAAAACCAATATTCCTCAACCAAGCTAATCCAGCAGGTGGCGCTGCCATCACTGA 1060
 QY 321 PheLyValaAspValaProGlyAspTyTyValaLeuValaAspHisAlaIlePheArgAlaPhe 340
 DB 1061 TTTAAGTGTGATGTCGCGGTGATTAATGCTGTTGACCATGCAATCTTCCGTCCCTTT 1120
 QY 341 AsnLyGlyValaLeuGlyTyTyLeuLyValaGlyGluGluAsnHisGlyIleTySer 360
 DB 1121 AACAAAGGGGCAATGGGCACTTAAGGTGAAGGTGAAGAAATCATAGATTAATTTCA 1180
 QY 361 HisLyGlnThrAspAlaValaTyTyLeuProGluGlyAlaProGlnAlaIleAspThrGln 380
 DB 1181 CACAAACCAACAGAGCTGTCTATCTCCAGAGGTTGCCCAACCAATTAATACCCA 1240
 QY 381 GluAlaProLyThrProAlaProAlaAsnLeuGlnGlnIleLyAlaGlyLyVala 400
 DB 1241 GAAGCACCCCAAAACACCTGACCTGCCAATTAACAAGCAGATTAAGCAGGTAAAGCA 1300
 QY 401 ThrTyTyAspSerAsnCyAlaAlaCyHisGlnProAspGlyLyGlyValaProAsnAla 420
 DB 1301 ACCATGATCTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1360
 QY 421 PheProProLeuAlaAsnSerAspTyTyLeuAsnAlaAspHisAlaArgAlaIleSerIle 440
 DB 1361 TTCACACCGCTTCCAACTGATCTGATCTGAAAGCCGACACGCTGTCGCCGACGATC 1420
 QY 441 ValaAlaAsnGlyLeuSerGlyLyValaIleThrValaAsnGlyAsnGlnTyTyGlySerValaMet 460
 DB 1421 GTGGCAAAATGATGTTGCTGTGTAAGATTACCGTCATATGCAATATGAAGAGCGTCAATG 1480
 QY 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValaIleThrTyThrLeuAsnSer 480
 DB 1481 CCTGGATTGCTCTAGGACCAACAGATGTCATGCTACCTTACCGTTAAACGC 1540
 QY 481 PheGlyAsnLyGlyGlyGlnLeuSerAlaAspAspValaAlaLyAlaLySerTyThrLy 500
 DB 1541 TTTGGTAAACAAAGCGGTCACTCACTGACAGACGATGTGGCAAAAGCCAAAACCAAG 1600
 QY 501 ProAsn 502
 DB 1601 CCAAC 1606

RESULT 4
 AX067454
 LOCUS AX067454 66986 bp DNA linear PAT 24-JAN-2001
 DEFINITION Sequence 29 from Patent WO0078968.
 ACCESSION AX067454
 VERSION AX067454.1 GI:12545074
 KEYWORDS
 SOURCE Moraxella catarrhalis
 ORGANISM Moraxella catarrhalis

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.
1
Lagace, R.E., Paterson, C. and Berg, K.L.
Nucleotide sequences of moraxella catarrhalis genome
Patent: WO 0078968-A 29 28-DEC-2000;
JOURNAL
Incyle Genomics, Inc. (US)
FEATURES
Location/Qualifiers
1..66986
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"

BASE COUNT 18889 a 13427 c 15112 g 19558 t
ORIGIN

Alignment Scores:

Score: 8.88e-165 Length: 66986
Percent Similarity: 2626.00 Matches: 502
Best Local Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-088-045-2 (1-502) x AX067454 (1-66986)

QY 1 MetSerLysProThrLeuIleLysThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20
DB 3742 ATGTCTAAGCCTACTTGTGATTAACAACCTTAATTTGGCTTAAGTGCATTTAGTCC 3801
QY 21 SerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLysSerSerThrValAspAla 40
DB 3802 AGTGGTTGTAGCAATCAACCGCAAGCCGCCCAAAAAGCAGCAGCGGTAGCGCT 3861
QY 41 AlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnIleGlnGlyLysLeuPro 60
DB 3862 GCCGCAGACAGCAAAATGCAATATGCTGCTCAAGAAATCAAGCAAGCGAGCTGCT 3921
QY 61 ValIleAspAlaIleValThrHisAlaProGluValProProValAspArgAspHis 80
DB 3922 GTCAATGATGCAATGTTTGCATGCAACGAAAGTTCCACCACTGTTGACCTGACACC 3981
QY 81 ProAlaLysValAlaValLysMetGlnThrValGluLysValMetArgLeuAlaAspGly 100
DB 3982 CCCGCCAAAGTGGTGTAAATAATGAAACCGTTGAAAAGTCATGCTGCGCAATGCG 4041
QY 101 ValGluThrGlnInPheThrPheGlyGlyGlnValProGlyGlnMetIleArgValArg 120
DB 4042 GTGGAATATCAAGTTTGGACATTTGGCGGTCAAGTTCCAGGCGAGATGATGCTGCT 4101
QY 121 GlnGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsn 140
DB 4102 GAAGCGACACATCGAAGTGCATTTCTAAACCCAGATTCACAAAATGCCCCATTAAT 4161
QY 141 ValAspPheHisAlaAlaThrGlyProGlyGlyAlaGlnAlaSerPheThrAlaPro 160
DB 4162 GTTGACTTTCAAGCTGCCACAGGCGCTGGCGGGGCGAAGCGCATTTATCCGACCG 4221
QY 161 GlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysLeuThrValIleArg 180
DB 4222 GGTCTATCACTCAACCTTTAGTTTAAAGCCTTACACCGTGTGATTCATCACTCT 4281
QY 181 AlaValAlaProValAlaGlyMetHisIleAlaAsnGlyMetThrGlyLysLeuValGlu 200
DB 4282 GGGGTGGCCCTGTTGGCATGCACTGCTTAATGCAATATGTTGATTTGGTTGA 4341
QY 201 ProLysGlnGlyLysProLysValAspLysGluThrValMetGlnIleAspPheThr 220
DB 4342 CCAAAAAGAGGCGCTGCGCAAAAGTAGATAAAGATATGTCATGCAAGCGCATTTAT 4401
QY 221 ThrLysGlyLysThrGlyGlnGlnGlnLysLeuGlnProPheAspMetGlnLysAlaIleArg 240
DB 4402 ACCAAAGGCAAAATATGCGCAACAGGTCTACAGCCCTTGTATATGAAAAAACCATTTGA 4461

QY 241 GlnAspAlaGluThrValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla 260
DB 4462 GAAGATGCTGAATATGTTGCTTAATATGTTGCGGGGGCGGTGACGTGAAATGCT 4521
QY 261 LeuLysAlaLysValGlyGlnThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu 280
DB 4522 CTAAAAGCGCAAGTTGGGAAACGTGCTTAATTTGGGTAAACGGGGCCGGAATTTG 4581
QY 281 ThrSerSerPheHisValIleGlyGlnIlePheAspLysValHisPheGlyGlyLys 300
DB 4582 ACATATATATTCATGATGATGTTGAGTAAATTTTGTATAGTTCACTTTGAGGGTGTAA 4641
QY 301 GlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGlu 320
DB 4642 GGTGAAAACCAATATCCAAACCAACCGCTAATCCAGAGGTGGCGGCTGCATCACTGA 4701
QY 321 PheLysValAspValProGlyAspThrValLeuValAspHisAlaIlePheArgAlaPhe 340
DB 4702 TTTAAGGTGATGTCGCCGGGTGATTAATGCTTGTGTTGACATGCTATTCCTGCTCT 4761
QY 341 AsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluGluAsnHisGluIleThrSer 360
DB 4762 AACCAAGGGCGATTTGGCATCTTAAGTGAAGGTGAAGAAATCATGACATTTATTC 4821
QY 361 HisLysGlnThrAspAlaValThrLeuProGluGlyAlaProGlnAlaIleAspThrGln 380
DB 4822 CACAAACAAACAGACGCTGTATCTGTCAGAGGGTGGCCCAAGCAATGATGATCCAA 4881
QY 381 GluAlaProLysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAla 400
DB 4882 GAAGCAACCAAAACCTGCACTGCACTGCACTGAACGCGCAACGCTGCGCGCCGCAATC 4941
QY 401 ThrThrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAla 420
DB 4942 ACCTATGATCTTAATGCTGCTGCTGTTGTCACCACTGATGTAAGGCTGCCAAACCT 5001
QY 421 PheProProLeuAlaAsnSerAspThrLeuAsnAlaAspHisAlaArgAlaAlaSerIle 440
DB 5002 TTCCACCGCTTGGCAACCTGCACTGCACTGCACTGAACGCGCAACGCTGCGCGCCGCAATC 5061
QY 441 ValAlaAsnGlyLysSerGlyLysIleThrValAsnGlyAsnGlnThrGluSerValMet 460
DB 5062 GTGGCAAAATGATGCTGTAAGATTAACCGTCATATGCAACATATGAAGAAACCTCATG 5121
QY 461 ProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrThrThrLeuAsnSer 480
DB 5122 CTTGCAATGCTGAGGACCAACAGATGCAATGATGATGATGATGATGATGATGATGATG 5181
QY 481 PheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysLysThrLys 500
DB 5182 TTTGTTAACAAGGCGTCAACTGATGTCAGATGTCAGATGTCAGATGTCAGATGTCAGATG 5241
QY 501 ProAsn 502
DB 5242 CCAAC 5247

RESULT 5
LOCUS AL646085/c 166050 bp DNA linear BCT 11-APR-2003
DEFINITION Ralstonia solanacearum GM11000 megaplasmid, complete sequence;
segment 10/11.
ACCESSION AL646085 AL646053
VERSION AL646085.1 GI:17431932
KEYWORDS
SOURCE Ralstonia solanacearum
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE
1 Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,
Arlat, M., Billault, A., Broctier, P., Camus, D. C., Cattelino, L.,
Chandler, M., Choisme, N., Claudel-Renaud, C., Cunne, S., Demange, N.,
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,


```

/translation="MARPAVQERGHMLKARQASSOPILLQSTLAPVDANIAAVEG
VLLAFVQFDVTIAGQKLAQVSDTELKQLEAEIASIOAOQALSAQOISSTEDYQ
AAARLLAONSLAQAORKEAOMLYDKGIARTLELAKQEDIDTQSOVIGADEI
ALRLKRAAPALRLVELTLENTROLLELRLAKLKTALLSPISGVLYVEPDGNDP
GKEELNAGAVTPKDVILITVDPSTAFLIKAWDEEDIRRVAGQAVRIYEGADSRAP
ACTRVSSQAARADGGGSGRGTAEFENQILFKPTPDAGRLRGSANMSLABERG
PTTLTPILAANVMSPAGRFVAVKRSASDTARMOETLEGRITLVDSVEVAGIADGEV
WPSAAGPDGDTLKQLLSGVNG"
gene
6828..7610
/gene="RSP1464"
/note="synonym: RS03080"
6828..7610
/gene="RSP1464"
/function="cell processes; transport of small molecules"
/note="Product confidence: probable"
Gene name confidence: hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by "FrameD"
/codon_start=1
/evidence=not experimental
/translation_table=11
/product="PROBABLE ATP-BINDING ABC TRANSPORTER PROTEIN"
/protein_id="CAD18615.1"
/db_xref="GI:17431938"
/db_xref="GOA:Q8XQ22"
/db_xref="SPTREMBL:Q8XQ22"
/translation="MAEPKAGNDRA PAGLLIELROVSRTPQVAGENVAAVRGVDLOVR
GSEVLAWVGPSSGSKSTMNLLGLDTPISGOYRLAGDTOTLSQESRLNQGIFG
VFOOFHLPALFALQAOPLFLYRGLSPASIDAPALVAAGLHRSRHAELSDG
ERQVLAAMAGOPRLILADEPTGLDSATGARVALALREVTATHTGALLMTIDLE
IARQLPROIEMDGLRCATQDPRPAAAPQPLPGRGTVP"
7612..8874
/gene="RSP1465"
/note="synonym: RS03079"
7612..8874
/gene="RSP1465"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence: probable"

Alignment Scores:
Pred. No.: 8.4e-89 Length: 166050
Score: 1495.50 Matches: 294
Percent Similarity: 69.92% Conservative: 71
Best Local Similarity: 56.32% Mismatches: 118
Query Match: 56.95% Indels: 39
DB: 1 Gaps: 9

US-10-088-045-2 (1-502) x AL646085 (1-166050)
QY 12 ILeCYaLaLeuSerAlaLeuMetLeuSerGlyCySerAsn----- 25
DB 50007 GTTTCGCTGTGTCAAGCTTCACGCTGCGCGCTCTACGTCGAGACACACAC 49948
QY 26 -----GlnAlaAspLysAlaIaIaGlnProLysSerSerThrValaAspAlaIaIaLys 43
DB 49947 ACCGACGACGCGGAGAAA-----AAATATAGCTGCTTCAACGCGTCCGCGT 49900
QY 44 ThrAlaAsnAla-----AspAsnAlaIaSer 52
DB 49899 CTGATCGCGCGCGCTGATTTTCGCTGACGCTGTGTACTGCTGACGACGCGCGGCGC 49840
QY 53 GlnGluHisGlnGlyGluLeuPro-----ValIleAsp 63
DB 49839 CAG-----GCAGCCAAAGCTTCCCGGCGATTTCGCTCCGCGCGCGCGACCCATCCGC 49786
QY 64 AlaIleValThrHisAlaProGluValProProProValaAspArgAspHisProAlaLys 83
DB 49785 GCGGCTCAGCAGTCCGCCCAACGTCGCGCGCGCGATCCACGCAATATCCGCGCAG 49726
QY 84 ValIleValLysMetGluThrValGluLysValMetArgLysLeuAlaAspGlyValGluTyr 103
DB 49725 GTTCATCTGGAGCTGGAACTGATCGAAGAAATGCGATCTCCGAAGGGGTCAAGTAC 49666

```

```

QY 104 GlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlyLysAsp 123
DB 49665 ACCTCTCGACCTTCGCGCGGACGAGTGCAGGACGACTTCATCCGCTGCGGCGAGGCGAT 49606
QY 124 ThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsnValaPhe 143
DB 49605 AGCGTGAGTCCACCTCAAGAACACCCGCGACACCAAGATCCGACACATCCACCTCG 49546
QY 144 HisAlaAlaThrGlyProGlyGlyGlnValaGluAlaSerPheThrAlaProGlyHisThr 163
DB 49545 CACGCGTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49486
QY 164 SerThrPheSerPheLysAlaLeuGlnProGlyLysLysValTyrHisCysAlaValaLys 183
DB 49485 TCGAGTTCACCTTCAAGGCGCTCAATAGAGGGGCTATGTCTTACACTGCGCGCACCGG 49426
QY 184 ProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGluProLysGlu 203
DB 49425 CCGGTCCGATGCATATATCGCAACGCGCATGTATGCGCTGATCTGTGACACCGCGGAA 49366
QY 204 GlyLeuProLysValaAspLysGlyTyrTyrValMetGlnGlyAspPheThrLysGly 223
DB 49365 GCGTCCCAAGTGAACACAGATCACTGATGACAGGGGATTTCTATACACCGGC 49306
QY 224 LysTyrGlyGlnGlyLeuGlnProPheAspMetGlyLysAlaIleArgGluAspAla 243
DB 49305 AAATACCGGAGAAAGGCCCGGACGCGTTCGACATGAAAGCCATGACGAGGCGCG 49246
QY 244 GluTyrValIleValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAla 263
DB 49245 ACCTACGTCCTTCAAGCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49186
QY 264 LysValGlyGluThrValArgLeuPheValGlyAsnGlyLysProAsnLeuThrSerSer 283
DB 49185 AAGACCGGAGACGCGTGGCATCTTCGTCGCAACGGGCGTCCAACTGCTGTGAC 49126
QY 284 PheHisValIleGlyGluIlePheAspLysValHisPheGluGlyLysGlyGluAsn 303
DB 49125 TTCACGATGTCGCGCATCTTCATCAGTGGCGCTACGAGGCGGACCAATGTGCAG 49066
QY 304 HisAsnIleGlnThrThrLeuIleProAlaGlyAlaIleThrGluPheLysVal 323
DB 49065 AAGATGTGCGACACCATCTGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49006
QY 324 AspValProGlyAspTyrValLeuValAsnHisAlaIlePheArgAlaPheAsnLysGly 343
DB 49005 AGGTGCGCGCGGACGTACGTGTGTGACACCTGATTTTCGCGCTTCAACAGGCG 48946
QY 344 AlaLeuGlyIleLeuLysValGluGluAsnHisGluIleTyrSerHisLysGln 363
DB 48945 GCGATGCGCATCTCAAAATGCAAGTCAGTCGCGGAGAACACCCCTGTATTTCGCGCAAGG 48886
QY 364 ThrAspAlaValTyrLeuProGluVala--ProGlnAlaIleAspThrGlnLysAla 382
DB 48885 CTGATGCGGTGTACTGTGCGACACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48826
QY 383 ProLysThrProAlaProAlaAsnLeu-----GlnGluGlnIleLysAlaLysAla 400
DB 48825 GCGCAGGCGTCCGATCCGCGACGCTGTGATGACGATGACGATGACGACGCGCGCGCG 48766
QY 401 ThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAla 420
DB 48765 CTGTTCCGCGGAGACCTGCTCGGTGTGCAACAGACCAAGCAACGAGGCTGCGCGCG 48706
QY 421 PheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaHisSerIle 440
DB 48705 TTTCCGCGGTGCGCAATGCGAATCTCTGCGCGCGGACCCGAAACGCGCGGTGAGATC 48646
QY 441 ValAlaAsnGlyLysSerGlyLysIleThrValaAsnGlyAsnGlnTyrGluSerValMet 460
DB 48645 GTGCTGACGCGCTGAACGCGCAAGTCAAGGTGAATGCGCAAGATTAAGATTCGCTGATG 48586

```


Oy		461	ProhibitinA---LseSerAsgGlnGlnIleLeuAlaenValIleTrrYrThrLeuAn	479
Dd		48585	CCCGCGAGACGCACGCTCAACGATGACGAGGTGCCAACACTCCTGACCTTAGTGCTTAAC	48522
Oy		480	SerPheGlyLeuNlySglYgLYGlnLeuSerAlaAspApVaAlaIysAlaIylSyrthr	499
Dd		48525	AGTTGGGACAACC CGGGGGCGCGGTCAGCGCGGAAGACGTC--AAAGACGATTGCTGCC	48465
Oy		500	LysPro 501	
Dd		48468	CAGCCC 48463	
RESULT 6				
AEO02512				
LOCUS				
DEFINITION		AE002512	Neisseria meningitidis serogroup B strain MCS8 section 154 of 206	
ACCESSION		AE002512	AE002098	
VERSION		AE002512.2	GI:7413469	
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

gene
CDS
/gene="NM01617"
complement (1209, .2078)
/note="similar to GB:M74072 SP:P25397 PID:149018 GB:U00966
PID:147680 PID:142062 percent identity: 68.42; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="cellulose resistance protein, putative"
/protein_id="AAFA1969.1"
/db_xref="GI:7226868"
/translation="MKERIVGSGELPCFGQPMVAKYENTPEVLLSGYSGESEGWCIL
NVLQDGVRAAPADSVASADGDCVAPAPQOVSVPKTDGAEIRLSLYCAADYPRH
KYGMSADHVSAAAPDTPAPAGRALDMDCCGGRNALFLGLKGEVTVADHNPALAYV
AELABEGINRTLEYDINAAALGSEPDYIATVLLFLMFQRPVDYIADQATTAAG
GNLIVLSAMDPAFPCCPMPFPFKREGLKDYRDMELVEYKEELGSMHAKDENGPI
RKFVTMLAKRPE"
2192, .2629
/gene="NM01618"
2192, .2629
/gene="NM01618"
2192, .2629
/note="similar to SP:P00647 PID:1208974 PID:147677
PID:147680 PID:142062 percent identity: 73.94; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ribonuclease H1"
/protein_id="AAFA1970.1"
/db_xref="GI:7226869"
/translation="MNQTVLYTDSACKGNPAGAGMGLMRYSHEKELFGSEKQTN
NRMETAVTEGLKRCCTVICTDSQYVNGKEMNTHCKRGMKTTAKQPKYOND
LKEKLDALVGHQVSWTVKKGHAGHAENRRADLANGAQFS"
3487, .3891
/gene="NM01619"
3487, .3891
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAFA1971.1"
/db_xref="GI:7226870"
/translation="MTAOPPCITCAQMEDVILQTPNLRVAVANDSSPAFCYIRK
HIAETDLSAERGELMEWYKVAAMROVFRPKINLASLGNVPHLHHIILARFEN
DASFPAPINAPVPRKHGWTLPQDWTDLKLL"
3931, .5115
/gene="NM01620"
3931, .5115
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAFA1972.1"
/db_xref="GI:7226871"
/translation="MTBPTSRRRFLKCTAAAGAGLLQAQCSATSVPPLPSSHSYVK
ARYVLTQPRRSDGNILRVASGSAEDTNRVNLTALTYNNGFVTVQAGSRRE
ORAGTGTORAPROFVAGSRVATPKVLMGLRGYGAARLPHIDASLGAARREHGT
LPFGSDVCAVOLALLAKGNMSPAGMAYSDBGRKAPGAFIMDAFIKGTQNRUTD
VPYIQRADVTEGLILMGNSLVLASLGTMPYIDIGLI.FLEDVGEQPIREMLMT
LYSLGILKQCAIVGVNFRMEKIRADVDPDSVSAVANNVSRKAI.PVLGPFPGHIA
DKITFIPGAHRIIRMNGNSGYVAFGEYPTLIDASALTLDLTPPDLPIFPESGVADI
SE"

gene
CDS
/gene="NM01621"
complement (5178, .5711)
/gene="NM01621"
complement (5178, .5711)

```

/gene="NM1621"
/Note="similar to PID:642472 SP:P52036 percent identity:
100.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glutathione peroxidase"
/protein_id="AAF41973.1"
/db_xref="GI:7226872"
/translation="MGIDYDFQKDAEGNVDLSGKGVLLIVNTATRCGLTPQYAL
QKLYAQYTAEGEILDFPCNOREQAPESGEIACQCMKFGTKFKIPKIEVNGANT
APLYVKEVKQKODGNHFKDFVLTALGKSRDGDIKMFTKFLVNRDSEVERF
ASVTPPEIEADRIALL"
/complement(5890..8145)
/gene="NM1622"
/complement(5890..8145)
/gene="NM1622"
/Note="similar to GP:2584871 percent identity: 78.28;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="nitric oxide reductase"
/protein_id="AAF41974.1"
/db_xref="GI:7226873"
/translation="MGQYKRLMTLPVAVLAVCFITIGKGVSEYKKAAPPYPROVYSAS
GVLTAKDIDLAGQSAWQTTGMEVGSILHGAYQAPMTALMLHRELAVMLDLTAQ
TYGKKEDEVSPQOAVLKTRLADENYRQVRKDSQVLSIDRVKAIESTLYHGVY
SDPRFOETREHAPAMKNTLPSQAEKLFNFPMTSASSTRPENETYNMPE
PLINNVPTTENMMSFTSYVLLMGILGMGLMFGFLTKIEEYVPEMDSIPALRPS
OKALKYVFLVALPVQVILGLTALHAYVGGFPGIDBEALGPEMSDMPALALRPS
HIQSAFMATGFLTRGLFLAPIVNGKPKFORAGVNTLYALPIVVGSGTAGNFA
LHLILPERFNFHOGIYLDGRWQLLVGLMLVLMRLCTVSAFKERGVKN
LAIIVASVWGVGAYPGLFYGEKSPVAVMEYRWVVALVWEGFEFAFAFAFV
FVNMGVRSTASTASTLAAAIKFLGSGVGTLLHLIFSGSTASMAIGAFSLAEVP
LVLRGEAYEHMSYHLSDMARKLRPLMCFVAVANMIGAVRFGFLINPISLEPI
OGLNLSAVAHAIKGVGFLALGFLVAVRLKPAQODDLMTMGFPLNLGLAM
IAISLIPVGAIOAYASITIGLMTYASRSEFLQHEILDITLKNVTRADLVITGALICAI
QATKIVFGDK"
/gene="NM1623"
/Note="similar to GB:M97926 SP:Q02219 PID:150277 percent
identity: 98.72; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="major anaerobically induced outer membrane
protein"
/protein_id="AAF41975.1"
/db_xref="GI:7226874"
/translation="MKRQALAAIASLPAALACGEPAAQAPETPAAAAEASSAQ
TAETPSGELPVIDATTHAPVPAIDIDYPAKVRKVEYKMTMDAGVRYRT
PGDYBPMIRVREGETVPESSNNSSVPHNVDPHATGCGGAAALFTAPGRST
RSFKALQPGILYHCAVAFVGMHIANGMISGLIVEPKGLPYDKEFLYVQSDFTKG
KKGAGLOPFMDKVAEQPEYVFNHGAIGADALAKAGETVRMYVNGGAPLV
SSEFHIGELFDYVVEGKLINENQSTIVPAGSAIVFEKVDIPESYTLVHSHIFRA
FKKALGOLKVGCAENPEIMTOKLSDTAVANGCAAPAAAPASAPASASEKSYV"
BASE COUNT      2477 a      2767 c      2385 g      2163 t
ORIGIN
Alignment Scores:
Pred. No.:      2,02e-83      Length:      9792
Score:          1394.00      Matches:      265
Percent Similarity: 79.74%      Conservative: 46
Best Local Similarity: 67.95%      Mismatches: 75
Query Match:     53.08%      Indels:      4
DB:              1          Gaps:      2
US-10-088-045-2 (1-502) x AE002512 (1-9792)
QY      1 MetSerLysProThrLeuIleLysThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20
DB      8498 TTACAAAGGAAATATTTGAAACGCCAACGCTTAGCTGCAATGATTCCTTCTTATTC 8557

```

```

QY      21 SerGlyCysSerAsnGlnAlaIleLysAlaIleGlnProLysSerSerThrValAspAla 40
DB      8558 GCATTAGCCGCTCGCGCGCGCGAAGCTCGCGCAAGCCCTCGCGAAGACCCCTCGCGCT 8617
QY      41 AlaAlaLysThrAla---AsnAlaAspAsnAlaAlaSerGlnGlnLHisGlnLysGlnLeu 59
DB      8618 GCGCGCGAAGCCGCAAGCTCGCGCGCGCAACACCGCGCGGAAACACCGCTCGCGCAATCG 8677
QY      60 ProValIleLeuAlaIleValThrHisAlaProGluValProProProValAspArgAsp 79
DB      8678 CCCGTTATCGATGCGGTTACCAACCCGCTCGGAGTCCCTCTCGCATCATGACCGCGAC 8737
QY      80 HisProAlaLysValValValLysMetGlnThrValGluLysValMetArgLeuAlaAsp 99
DB      8738 TACCCGCGCAAGTCCGCGTTAAATGGAACCGTCGAAAAAACATGACCATGGAAGAC 8797
QY      100 GlyValGluThrGlnPheThrThrPheGlyGlyGlnValProGlyGlnMetIleArgVal 119
DB      8798 GGTGTGGAATACCGGTACTGACATTTGACGCGGACGTTCCGCGCGGTATGATCCGCGTA 8857
QY      120 ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139
DB      8858 CCGGAGGCGATACCGGTGAAGTGAATTTTCCAAATCTTCTTACCGTTCGCGAC 8917
QY      140 AsnValAspPheHisAlaAlaThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAla 159
DB      8918 AACGTGCACTTCCACGCGGCTACCGGCGGCGGCGGCGGCGGCGCAACTTTACCGCT 8977
QY      160 ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysThrValThrHis 179
DB      8978 CCGGCGCGCTACTTCACTTCACTTCAAGCCCTGCAACCGGCTGTACTACCAAC 9037
QY      180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetThrGlyLeuIleLeuVal 199
DB      9038 TCGCGCGTCCGACCGGTGATGCAATCGCCCAACGATATATGCGTGTGATTTTGTC 9097
QY      200 GluProLysGluGlyLysProLysValAspLysGluThrThrValMetGlnLysAspPhe 219
DB      9098 GAGCTTAAAGAGCGCTCGCGAAGTGAATGAAGTTCTATCATGCTCAAGCGCATTC 9157
QY      220 TyrThrLysGlyLysThrGlyGlnGlnLysLeuGlnProPheAspMetCysLysAlaIle 239
DB      9158 TACACCAAGGCAAGGCAAGGCGCGCAAGTCTGCAACCGTGTGATATGCAAGCGCTT 9217
QY      240 ArgGluAspAlaGluThrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsn 259
DB      9218 GCGGAACGCTGAATAGCGTATTCACGCTCACGTAGGTGCTATCGCGCGCATAC 9277
QY      260 AlaLeuLysAlaLysValGlyLysThrValArgLeuPheValGlyLysAsnGlyProAsn 279
DB      9278 GCGCTGAAGGCAAGGCGGCAAGCTGATGATGATGATGATGATGATGATGATGATGATG 9337
QY      280 LeuThrSerSerPheHisValIleGlyLysIlePheAspLysValHisPheGluGlyGly 299
DB      9338 TTGGTATCTTCTTCTTCAAGTCACTCGCGCAAACTTTCGCAAAAGTTATTTGAAAGCGCG 9397
QY      300 LysGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyAlaAlaIleThr 319
DB      9398 AACTGATTTAAGAAACGTATCAAGACACATCGTTCGCGCGGCGCTGCTGCTGCTGCTG 9457
QY      320 GluPheLysValAspValProGlyAspThrValLysValAspHisAlaIlePheArgAla 339
DB      9458 GAATTCAAAGTGAATCCCGGCGAGTCACTTGTGTTGACCATCTATCTTCCGCGCA 9517
QY      340 PheAsnLysGlyAlaLeuGlyLysLeuLysValGluGlyGluLysAsnHisGluIleThr 359
DB      9518 TTCAACAAAGGCGCACTGGTCAATTTGAAGTAGAAGTGCAGAAACCTTGAATATCATG 9577
QY      360 SerHisLysGlnThrAspAlaValLysThrLeuProGluGlyAlaProGlnAlaIleAspThr 379
DB      9578 ACTCAAAATGATGATACCGCTTACCGCGGTAACGGTGAAGCTCTGCTGCTGCTGCTGCTG 9628

```

OY 380 GlnGluAlaProlyThrProAlaProAla 389
 Db 9629 GCTTCCGCTCCCGAGCTTGTGCCCGCA 9658
 RESULT 7
 NMA622491 326301 bp DNA linear BCT 02-SEP-2002
 LOCUS Neisseria meningitidis serogroup A strain 22491 complete genome;
 DEFINITION segment 6/7.
 ACCESSION AL162757 AL157959
 VERSION AL162757.2 GI:7380371
 KEYWORDS
 SOURCE Neisseria meningitidis 22491
 ORGANISM Neisseria meningitidis 22491
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE 1 (bases 1 to 326301)
 AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
 Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
 Davies,R.M., Davis,P., Devlin,K., Felwell,T., Hamlin,N.,
 Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
 Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
 Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
 TITLE Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491
 JOURNAL Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE 2 (bases 1 to 326301)
 AUTHORS Parkhill,J.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
 COMMENT Details of N. meningitidis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/N_meningitidis/).
 FEATURES
 source 1. 326301
 /organism="Neisseria meningitidis 22491"
 /mol_type="genomic DNA"
 /strain="22491"
 /db_xref="taxon:122587"
 /note="serogroup: A"
 196. .311
 /note="REP 2; hmms hit to HMM REP 2 (1 - 109), score:
 87.43"
 434. .453
 /note=">= 90% match to ATCCNNNNNNNGCAAT"
 /label="dRS3"
 498. .560
 /note="Correia element; hmms hit to HMM Correia (1 - 62),
 score: 78.71"
 /label="Correia"
 562. .604
 /note="Correia element; hmms hit to HMM Correia (114 -
 156), score: 51.92"
 /label="Correia"
 complement (667. .676)
 /note="Core DNA uptake sequence: gccgcctgaa"
 /label=DUS
 complement (668. .1813)
 /gene="dape"
 complement (668. .1813)
 /gene="dape"
 complement (668. .1813)
 /gene="dape"
 /EC_number="3.5.1.18"
 /note="NMA1730, dape, succinyl-diaminopimelate
 desuccinylase, len: 381aa; similar to many eg. SW:P24176
 (DAPE_ECOLI) dape, succinyl-diaminopimelate desuccinylase
 from Escherichia coli (375 aa) fasta scores; E(): 0, 56.9%
 identity in 371 aa overlap. Contains Pfam match to entry
 PF01546 Peptidase_M20."

/codon_start=1
 /transl_table=11
 /product="succinyl-diaminopimelate desuccinylase"
 /protein_id="CAB84958.1"
 /db_xref="GI:7380372"
 /db_xref="SPTREMBL:Q9JTL0"
 /translation="MTETQSLELAKALISRPSTPDDRCCKLVERLYKIGFAABEL
 HPGDTRKNIWLRGTQVYVCFAGHTDVPTGVEKWDSPFETRDGLYCGAADM
 KTIACFPVACERFVAEHPDHOGSIALITSPEDDADGCTTKVVDYKARGLIDYC
 IYGEPRADKXGDMTKNGRSGISGNLTVKCGKSHIADPHLAINVHTAPALLETQ
 EYWDGKNKFPPTSTFOISVINGCTGATNVIPEELVKRNFRTSTSTEGLKORVHAI
 LDKHVOYDLQWSSCGQPLTQAGKLTVAARAIAETGIEALELSTGTSGRPIKA
 IAKELIELEGPNAWTHQINENVRDLDPRLSAVYEGIIARLAGNAV"
 complement (836. .1786)
 /gene="dape"
 /note="Pfam match to entry PF01546 Peptidase_M20, score
 252.50, E-value 5.7e-72"
 complement (1863. .1872)
 /note="Core DNA uptake sequence: gccgcctgaa"
 /label=DUS
 complement (1927. .2547)
 /gene="NMA1731"
 complement (1927. .2547)
 /gene="NMA1731"
 /note="NMA1731, conserved hypothetical protein, len:
 206aa; similar to SW:P44013 (Y552 HAEIN) hypothetical
 protein from Haemophilus influenzae (207 aa) fasta scores;
 E(): 0, 59.6% identity in 203 aa overlap."
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="CAB84959.1"
 /db_xref="GI:7380373"
 /db_xref="SPTREMBL:Q9JQZ3"
 /translation="WLTPKSCDLENIIPFPQSOLKKYPPESIPQIKADYKENWQIQWQ
 LIQVAAELGAFAPAPHIERMCGVQVAHFPAKYQYKNSAIIISILNRRLSV
 SLDWMCYKADVPALPDYRNLDNFDTEKASFPMMGABSEYDYDTVAQONSBR
 KRONEDPFCIGKHTERDLDGRODVAKVIAETVEDLLPLYERCHGK"
 complement (2551. .2558)
 complement (2570. .2579)
 /note="Core DNA uptake sequence: gccgcctgaa"
 /label=DUS
 complement (2581. .3090)
 /gene="NMA1732"
 complement (2581. .3090)
 /gene="NMA1732"
 /note="NMA1732, conserved hypothetical protein, len:
 169aa; similar to SW:P43971 (Y243 HAEIN) hypothetical
 protein from Haemophilus influenzae (172 aa) fasta scores;
 E(): 0, 49.7% identity in 169 aa overlap."
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="CAB84960.1"
 /db_xref="GI:7380374"
 /db_xref="SPTREMBL:Q9JTK9"
 /translation="WNPETOSVTPABEITMLVACHGKVRPFCQVAMLSYIAENG
 NQVIVQTRQUTRYNVNAPLPHHDEBENFPPLIQVAPQOQSEYDELRLHVSJH
 WDVAVERPAKLADNAVTPDAEAFRRFVAGVDVHLAIEPLFDWGMKPIPEKLTETG
 EIMARRRR"
 complement (3185. .3736)
 /gene="NMA1733"
 complement (3185. .3736)
 /gene="NMA1733"
 /note="NMA1733, H.8 outer membrane protein, len: 183aa;
 strongly similar to SW:P07212 (H8 NEIME) H.8 outer
 membrane protein from Neisseria meningitidis (183 aa)
 fasta scores; E(): 0, 96.7% identity in 183 aa overlap.
 Contains Pfam match to entry PF00127 copper-bind, Copper
 binding proteins, plastocyanin/azurin family and Prosite
 match to PS00196 Type-1 copper (blue) proteins signature."
 /codon_start=1
 /transl_table=11

QY 381 GluAlaProLyThrProAlaProAla 389
 Db 149109 TCCGCTCCCGACCTCTGCTCTGCA 149135

RESULT 8
 AX044033 349980 bp DNA linear PAT 24-NOV-2000
 LOCUS Sequence 112 from Patent WO0066791.
 DEFINITION AX044033
 ACCESSION AX044033
 VERSION AK044033.1 GI:11342917

KEYWORDS
 SOURCE Neisseria meningitidis
 ORGANISM Neisseria meningitidis
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.

REFERENCE
 AUTHORS Pizzo, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C., Masiarant, V., Galeotti, C., Mora, M., Ratti, G., Scarfelli, M., Scariato, V., Rappunli, R., Frazer, C. M. and Grandi, G.
 TITLE Neisseria genomic sequences and methods of their use
 JOURNAL Patent: WO 0066791-A 112 09-NOV-2000;
 CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES
 source
 1. 349980
 Location/Qualifiers
 /organism="Neisseria meningitidis"
 /mol_type="genomic DNA"
 /db_xref="taxon:487"
 /note="sequence too long, cut in 8 pieces.-seq 1: 1 to 349980 349980 bases-seq 108: 300001 to 649980 349980 bases-seq 109: 600001 to 949980 349980 bases-seq 110: 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to 222325 172325 bases"

BASE COUNT 87189 a 93501 c 84627 g 84663 t

ALIGNMENT SCORES:
 Pred. No.: 1 19e-81 length: 349980
 Score: 1394.00 matches: 265
 Percent Similarity: 79.74% conservative: 46
 Best Local Similarity: 67.95% mismatches: 75
 Query Match: 53.08% indels: 4
 DB: 6 gaps: 2

US-10-088-045-2 (1-502) x AX044033 (1-349980)

QY 1 MetSerLybProThrLeuLeuLyThrLeuLeuLeuLeuLeuLeuLeu 20
 Db 187900 TTCACAAAGGAAATATATGAAAGCGCAAGCCTTAGCTGATGATGCTTCTTATTC 187959

QY 21 SerGlyCySerSerGlnAlaLeuAlaLeuAlaLeuAlaLeuAlaLeu 40
 Db 187960 GCATTAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 188019

QY 41 AlaAlaLeuThrAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeu 59
 Db 188020 GCCCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 188079

QY 60 ProValIleAlaAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeu 79
 Db 188080 CCGGTTATCGATCGGTTATCGATCGATCGATCGATCGATCGATCGATCG 188139

QY 80 HisProAlaLeuValValValValValValValValValValValVal 99
 Db 188140 TACCCCGCAAGTCCGCGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCA 188199

QY 100 GlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 119
 Db 188200 GGTGTGAATACCGCTACTGACATTTGACGGGCGGCGGCGGCGGCGGCT 188259

QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLybMetProHis 139

Db 188260 CCGCAAGCGGATACGGTGAAGTGAATTTCCAAATCTCTTCCATCCGTCGCGAC 188319

QY 140 AsnValAspPheHisAlaAlaThrGlyProGlyGlyValAlaGluAlaSerPheThrAla 159
 Db 188320 AACGTGACCTTCCACCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCAACTTTACCGCT 188379

QY 160 ProGlyHisLeuSerThrPheSerPheLybAlaLeuGlnProGlyLeuThrValTyHis 179
 Db 188380 CCGGCGCGTACTTCCATCTTACCTTCCAAAGCGCTTCCAAAGCGGCTTGTACATCTACCC 188439

QY 180 CysAlaValAlaProValGlyLeuHisIleAlaLeuGlyMetTyGlyLeuVal 199
 Db 188440 TCGCGCGTGCACCGGTCGGTATGACATCGCCAAAGGATATGACGTGATTTGGTC 188499

QY 200 GluProLybGlyLeuProLybValAspLybGlyTyTyValMetGlnLybPhe 219
 Db 188500 GAGCCTAAAGAGGCGCTCCGCAAGTGAATGAAGTTCATCATGTCACAGGCACTTC 188559

QY 220 TyrThrLybGlyLeuTyGlyGlyGlyLeuGlnProPheAspMetGlyValAlaIle 239
 Db 188560 TACACCAAGGCAAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 188619

QY 240 ArgGluAspAlaGlyTyValValPheAsnGlySerValGlyAlaLeuThrGlyGlu 259
 Db 188620 GCCCAAGCGCTGAATACGTCGATTAACAGGTCACGTCGTCGTCATGCGCGCGCATAC 188679

QY 260 AlaLeuLybAlaLeuValGlyValValValValValValValValValVal 279
 Db 188680 GCGCGCAAGGCAAAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 188739

QY 280 LeuThrSerSerPheHisValIleGlyGlyLeuPheAspLybValHisPheGlyGly 299
 Db 188740 TTGGTATCTTCTTCCATCTTACGTCGTCGCGCAATCTTCCGCAAGTTATGTAAGCGGC 188799

QY 300 LybGlyGluAsnHisAsnIleGlnThrThrLeuLeuProAlaGlyValAlaIleThr 319
 Db 188800 AAATGATTAACGAAGAGCTACGAAGCAACATCGTTCGCGCGCGCTCCATCGTC 188859

QY 320 GluPheLybValAspValProGlyAspTyValValLeuValAspHisAlaIlePheThrAla 339
 Db 188860 GAATTCAAAGTGCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 188919

QY 340 PheAsnLybGlyAlaLeuGlyTyLeuLybValGlyGlyGlyAlaHisIleGlyTy 359
 Db 188920 TTCACCAAGGCGCGCGCTGCTAATGAAAGTGAAGTGCAGAAACCTGAATCATG 188979

QY 360 SerHisLybGlnThrAspAlaValTyTyLeuProGlyAlaProGlnAlaIleAspThr 379
 Db 188980 ACTCAAAATTAAGTATACCGCTTACCGCGGTAAGGTCAGTCCTGCT----- 189030

QY 380 GlnGluAlaProLybThrProAlaProAla 389
 Db 189031 GCTTCGCTCCGCGAGCTTTCGCGCGCA 189060

RESULT 9
 LOCUS NGOPANI 2069 bp DNA linear BCT 26-APR-1993
 DEFINITION Neisseria gonorrhoeae major anaerobically induced outer membrane protein (Pan1) gene, complete cds.
 ACCESSION M97926
 VERSION M97926.1 GI:150276
 KEYWORDS anaerobically induced protein; major outer membrane protein.
 SOURCE Neisseria gonorrhoeae
 ORGANISM Neisseria gonorrhoeae
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.

REFERENCE
 AUTHORS Hoehn, G. T. and Clark, V. L.
 TITLE Isolation and nucleotide sequence of the gene (an1) encoding the major anaerobically induced outer membrane protein of Neisseria gonorrhoeae
 JOURNAL Infect. Immun. 60 (11), 4695-4703 (1992)

MEDLINE 93014187
 PUBMED 1383156
 COMMENT Original source text: Neisseria gonorrhoeae (strain R10) DNA.
 FEATURES Location/Qualifiers
 source 1.2069
 /organism="Neisseria gonorrhoeae"
 /mol_type="genomic DNA"
 /strain="R10"
 /db_xref="taxon:485"
 1..1265
 /gene="Pan1"
 <1..50
 /gene="Pan1"
 /note="putative"
 76..79
 /gene="Pan1"
 /standard_name="Shine-Delgarno sequence"
 87..1265
 /gene="Pan1"
 /codon_start=1
 /transl_table=11
 /product="major anaerobically induced outer membrane protein"
 /protein_id="AA025462.1"
 /db_xref="GI:150277"
 /translation="MKRQALAAAMIASLPALAAAGSGROAQAPEETPAASAEASAAQ
 AATPAGELPVIDATVTHAPVPAIDIDYPAKRVKMEVETKMDQDVEEYWT
 PFGDVGPMIRVREGDIVEVERSNPSSDPVPHNVPHATGGGAAATPTAPGRTST
 PSFKALQPGELTYTHCAVAPVGHNIANGKGLIVPERKSLPRVDKEFTYVQDPPTTKG
 KKGAQGLQPDMDKVAEQPERYVFNHNGSIAQDNALAKAKGERTVRNKGSGENLV
 SFFHVGELPDKRYVEGKLIENVOSTIVPAGSAIYEFKVDIPGSYTLVDHSIFRA
 FNKGALGOLKVEGAENPEIMTOKLSDTAVAGSAAAPAPASAPASAPASASERSVY
 " repeat_region 1358..1417
 /standard_name="pan1"
 BASE COUNT 526 a 611 c 511 g 421 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.68e-84 Length: 2069
 Score: 1392.00 Matches: 263
 Percent Similarity: 80.26% Conservative: 50
 Best Local Similarity: 67.44% Mismatches: 75
 Query Match: 53.01% Indels: 2
 DB: 1 Gaps: 2
 US-10-088-045-2 (1-502) x NC00PANI (1-2069)
 QY 1 MetSerLysProThrLeuIleuLysThrThrLeuIleuLysLeuSerAlaLeuSerAlaLeuMetLeu 20
 Db 69 TTACAAAGGAAAGCACTTATGAAGCCGCAAGCATTCGCGCAATGATTCCTTCTTATTC 128
 QY 21 SerGlyCysSerAsnGlnAlaAspLysAlaIleGlnProLysSerSerThrValAspAla 40
 Db 129 GCATTGCGCGCTCGCGTGGCGAACAACGCTGCTCAAGCCCTGCGCAAAACCCCTGCGCGCT 188
 QY 41 AlaIleuLysThrAla---AsnAlaAspAsnAlaIleSerGlnGlnIleGlnIleGlnVal 59
 Db 189 TCCGGAAGAGCCGCAAGTTCGCGCGCAACGCTACCGGAAACGCTTCGAGCGCAATCG 248
 QY 60 ProValIleAspAlaIleValThrHisAlaProGluValProProProValAspArgAsp 79
 Db 249 CCGGTCATGATGCGGTGACCAACCGCTCCGGAAGTACCTCCGCAATCGACCGCGAC 308
 QY 80 HisProAlaLysValValValLysMetGluThrValGluLysValMetArgLeuAlaAsp 99
 Db 309 TATCTCTGCAAAAGTACGCTAAAGAAACGCTGAAACCACTGAAATGACGAC 368
 QY 100 GluValIleGluTyrGlnPheThrPhePheGlyGlnValProGlyGlnMetIleArgVal 119
 Db 369 GGGGTGAAATACCGGCTGACATTTGACGGGACGTTCCGGGCGGTATGATCGCGGTA 428
 QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139

Db 429 CGCGAAGCGATACCGTTGAAGTCGAATTCCTCCAAATCTTCTTACCGCTTCGAC 488
 QY 140 AsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluAlaSerPheThrAla 159
 Db 489 AACGTGACTTCCACCGCCGCAACCGCTCAGGGCGGGGTGACGCCGCACTTTCACGCC 548
 QY 160 ProGlyHisIleSerSerThrPheSerPheLysAlaLeuGlnProGlyLysThrValTyrHis 179
 Db 549 CCGGCGCGCACTTCACATTCAGCTTCAAAAGCCCTGCAACCGGCGCTGACATTCACAC 608
 QY 180 CysAlaValAlaProValIleGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuVal 199
 Db 609 TCGCGCGTCCGCGCGCTGATGACATCCGCAACGATATGATGATGATGATGATGATGATG 668
 QY 200 GluProLysGluGlyLeuProLysValAspLysGluTyrTyrValMetGlnIleAspPhe 219
 Db 669 GAGCTTAAGAAAGGCTCCGAAAGTGAATGAAGGTTCTATCATGTCATGTCACAAACCTTC 728
 QY 220 TyrThrLysGlyLysTyrGlyGluGlnIleLeuGlnProPheAspMetGluLysAlaIle 239
 Db 729 TACACCAAGCAAGCAAAAGGCGCGCAAGCGCTGCAACCGTTGATGATGATGATGATGATG 788
 QY 240 ArgGluAspAlaGluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsn 259
 Db 789 GCCGAACAGCCGGAATATACGTCGATTCACCGCTCATGATGATGATGATGATGATGATG 848
 QY 260 AlaIleuLysAlaLysValGlyGluThrValArgLeuPheValGlyValAsnGlyGlyProAsn 279
 Db 849 GCCCTGAAGGCCAAAGCAAGCGCAAAACCGTGGCTATGATGATGATGATGATGATGATG 908
 QY 280 LeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGlyGly 299
 Db 909 TTGGTGCTCTCTCTTCCAGTATGATGCGGCAATCTTCGCAAAAGTTATGTTAAAGCGCG 968
 QY 300 LysGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaIleThr 319
 Db 969 AAACGTATTAACGAAGACGTAACGAAGACGATCGCTGCGCGCGCTTCCATCGTC 1028
 QY 320 GluPheLysValAspValProGlyAspLysPheValLeuValAspHisAlaIlePheArgAla 339
 Db 1029 GAATTCAAAGTCGATCCCGGCGCTACCTTGGTGACACTTCCATCTTCGCGCA 1088
 QY 340 PheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluGluAsnHisGluIleTyr 359
 Db 1089 TTCAACAAAGCGCTTGGGCGAATTCGAAGTAGAGGCTGCCGAACCTGAAATCATG 1148
 QY 360 SerHisLysGlnThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThr 379
 Db 1149 ACTCAAAATTCAGATGATACCGCTTACCGCGGACGCGGCGCTTCTGCC---CCTGCT 1205
 QY 380 GlnGluAlaProLysThrProAlaProAla 389
 Db 1206 GCTTCGCAACCGGCTGCTTCTGCCCCGCA 1235
 RESULT 10
 AB076606 3675 bp DNA linear BCT 15-JUN-2002
 LOCUS Hypomicrobium denitrificans nirK gene for nitrite reductase,
 complete cds.
 ACCESSION AB076606
 VERSION AB076606.1 GI:21623659
 SOURCE
 ORGANISM Hypomicrobium denitrificans
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Hypomicrobiales; Hypomicrobium.
 REFERENCE
 AUTHORS Karasaka, K., Fukui, A., Kabayashi, M., Yamaguchi, K. and Suzuki, S.
 TITLE Cloning and expression of copper-containing nitrite reductase from
 Hypomicrobium denitrificans
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3675)

AUTHORS
Kataoka, K. and Suzuki, S.

TITLE
Direct Submission
Submitted (19-DEC-2001) Kunihide Kataoka, Kanazawa University,
Department of Chemistry, Faculty of Science, Kakuma, Kanazawa,
Ishikawa 920-1192, Japan
(E-mail) kataoka@chem.kanazawa-u.ac.jp, Tel: 81-76-264-5683,
Fax: 81-76-264-5742)

FEATURES
Location/Qualifiers
1. .3675
/organism="Hyphomicrobium denitrificans"
/mol_type="genomic DNA"
/strain="A3151"
/db_xref="taxon:53399"
36. .1496
/gene="nirK"
36. .1496
/gene="nirK"
/codon_start=1
/transl_table=1
/product="nitrile reductase"
/protein_id="BAC00912.1"
/db_xref="GI:21623660"
/translation="MFRFGAVGMAALLSVPAHADQHQMAQKAPADAPAM
KDKSKSEBKTDPTAGAAAPVFTLRGIABRMVYIGVGIDIKINPTLVHEGE
TVQVNLVNBEGADIVVDQYARSAIVGKVASSTFVSASVKGSEYCSIAHRO
AGMEGNIQVLPGRABEMKSGADITRDPDLPGLPQAKTVRIDLETVYVQGLDD
NTYTYWTFNGKAGPFLFRVGVDTVELKHKHKGDSIMVSHVDFHAGATPGGAATFO
TQDGEIVTFKALIPGIYVHCATPSVPHITNGMYGLIVBEGGLPOVDERPFRF
OEIYTVKSPCTSGBOEMYEKINKEPFLNGSVGLTSHRPYASVGEYRFRF
GUGGNFTSSPHVIGETIPHVSVLSGSVSPGLGVTVSPPGATTVDFKIDRGY
ILVDHALSLREHLVGLVNDGPKNDSIMHEGPQK"

gene
36. .1496
/gene="nirK"
/codon_start=1
/transl_table=1
/product="nitrile reductase"
/protein_id="BAC00912.1"
/db_xref="GI:21623660"

CDS
36. .1496
/gene="nirK"
/codon_start=1
/transl_table=1
/product="nitrile reductase"
/protein_id="BAC00912.1"
/db_xref="GI:21623660"

sig_peptide
36. .152
/gene="nirK"
/evidence="experimental"

BASE COUNT 817 a 1122 c 1032 g 704 t
ORIGIN
1522 1552

Alignment Scores:

Pred. No.: 2,366-43 **Length:** 3675
Score: 787.00 **Matches:** 177
Percent Similarity: 50.544 **Conservative:** 56
Best Local Similarity: 38.394 **Mismatches:** 135
Query Match: 29.974 **Indels:** 94
DB: 1 **Gaps:** 8

US-10-088-045-2 (1-502) x AB076606 (1-3675)

QY 55 HsGlnGlyLeuProValIleAspAlaIleValThrHisAlaPro---GluValPro 73
Db 552 AATCGCGCGGAGATGAATCGTCCGGAGCGGACATCAGCGGTGACCCCGCGATCGCA 611
QY 74 ProProValAspArgAspHisProAlaValValValValMetClnThrValGluVal 93
Db 612 GGCCTCATCGCCCGCGTCAAGCAAGCGTCCGATCATCAAACTGTCGAGGTC 671
QY 94 ValMetArgLeuAlaAspGlyValGluValGlnPheThrPheGlyGlyGlnValPro 113
Db 672 AAGGACAGCTGACGACCAACAGCACTACGACTGACGCTTCAACGCAAGTTCCG 731
QY 114 GlyIleMetIleArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHisPro 133
Db 732 GGTCCCTTCCTTCGCGTCCGCGTGGACACGCGTGCATTTGAAAATATCAAA 791
QY 134 AspSerIleMetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyAla 153
Db 792 GACAGCTTGATGATCTACCTCGGACTTCCAGCGGCAACCGGACCTGCGCGAGCT 851
QY 154 GluAlaSerPheThrAlaProGlyHisIleThrSerPheSerPheValIleGlnPro 173
Db 852 GCATTACCGCAACCGATCCCGCGCAAGAAACCGTGTGACGTTCAAGCGCTGATTCCC 911
QY 174 GlyLeuValValValValValValValValValValValValValValValValVal 193

Db 912 GGCATCTACGCTATACATCGCCGACGCGGACGCTGCGGACCGACATCCAAACGCGATG 971
QY 194 TyrGlyLeuIleLeuValGluProGlyGlyGlyLeuProValAspValGlyVal 213
Db 972 TACGGATCTTCCTCTGATGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1031
QY 214 ValMetGlnGlyAspPheThrValGlyValGlyValGlyValGlyValGlyValGlyVal 233
Db 1032 GTCATTCAGGCGGAGATCTATACGATCAAAATCGCTTCCGACCTCCGCGGACGAGAAAG 1091
QY 234 AspMetGluValAlaIleArgGluAspAlaGluValValValPheAsnGlySerValGly 253
Db 1092 GACTACGAAAGCTGATCAACGAAAGCGGAGTACTTCTTCAACGCGTCCGCTCGGCT 1151
QY 254 AlaLeuThrGlyGluAsnAlaLeuValValValValValValValValValValValVal 273
Db 1152 TCGCTTACCGCGTACCGATCCGCTTATGCGACGCGGCGGAAACGCGGCGATCTTCTTC 1211
QY 274 GlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlyValPheAspVal 293
Db 1212 GCGCTCGCGGACCGAATCTTCACTGATCTTCCACGTCATCGTGAAATCTTCCATCAT 1271
QY 294 ValHisPheGlyGly-----LysGlyGluAsnHisAsnIleGlnThrThrLeuIle 311
Db 1272 GTCTATTCACTCGGAAAGCTGTTTCCCTCCGCTGATCGCGGCGGCGGCTTCTCGTT 1331
QY 312 ProAlaGlyValAlaAlaIleThrValPheValAspValProGlyAspValValLeu 331
Db 1332 CTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1391
QY 332 ValAspHisAlaIlePheArgAlaPheAsnValGlyAlaLeuGlyIleLeuValValGlu 351
Db 1392 GTTCATCATGCTCTCTCTCTGCG---CTGAGACATGCGCTCTGCGCTTCTGAACTGAC 1448
QY 352 GlyGluGluAsn-----HisGlu----- 357
Db 1449 GGTCCGAAAGACGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1508
QY 357 ----- 357
Db 1509 TCCGAAGGTAAACCGAGGCTCAGCGCTTGAAGCTGAGCGCTCCGTAATTTATTTGAGATA 1568
QY 358 ----- 1568
Db 1569 CCGGACGCTTAAGACAGCTTCCGGAACCGGAAGAAAGAGCGGCGGCGGCGGCGGCGGCGG 1628
QY 359 TyrSerHisLysGlnThrAsp-----AlaValVal 368
Db 1629 CAGCTGCATGCGGAAACAAAGACGCGGCTGCGTGAATCGCGTCTCCGCAACG 1688
QY 369 LeuProGluValAlaProGlnAlaIleAspThrGlnGluAlaProValThrProAlaPro 388
Db 1689 ATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1748
QY 389 AlaAsnLeuGlnGluGlnIleValAlaGlyValValValValValValValValValValVal 408
Db 1749 GCGACGCG---CGTACCGCGGATACGCGGCGGATACGCGGCGGAT----- 1798
QY 409 CysHisGlnProAspGlyValValPro----- 418
Db 1799 GCTCATCAACCGATGCTCGATACCCAGGATGAACCGGCGGCTTGGCTCAGCGG 1858
QY 419 ----- 419
Db 1859 CGACGCGCAACAGACGCGGCGGATGAAGAGAGATCAACGAGGATTCAGCGATGAGCC 1918
QY 426 Asn 426
Db 1919 AAC 1921
RESULT 11
HMA278286

LOCUS HMA278286 2506 bp DNA linear BCT 27-JUN-2001
 DEFINITION Haloarcula marismortui nirK gene for copper-containing dissimilatory nitrite reductase and pcn gene for proliferating cell nuclear antigen.
 ACCESSION AJ278286
 VERSION AJ278286.1 GI:8249445
 KEYWORDS copper-containing dissimilatory nitrite reductase; nirK gene; pcn gene; proliferating cell nuclear antigen.
 SOURCE Haloarcula marismortui
 ORGANISM Archaea; Euryarchaeota; Halobacteria; Halobacteriales; Halobacteriaceae; Haloarcula.
 REFERENCE 1 Ichiki, H., Tanaka, Y., Mochizuki, K., Yoshimatsu, K., Sakurai, T. and Fujiwara, T.
 TITLE Purification, characterization, and genetic analysis of Cu-containing dissimilatory nitrite reductase from a denitrifying halophilic archaeon, Haloarcula marismortui
 JOURNAL J. Bacteriol. 183 (14), 4149-4156 (2001)
 MEDLINE 21311739
 PUBMED 11418554
 REFERENCE 2 (bases 1 to 2506)
 AUTHORS Fujiwara, T.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAY-2000) Fujiwara T., Department of Biology and Geosciences, Faculty of Science, Shizuoka University, 836 Oya, Shizuoka 422-8529, JAPAN
 FEATURES
 source
 1. 2506
 /organism="Haloarcula marismortui"
 /mol_type="genomic DNA"
 /db_xref="taxon:2238"
 269..1449
 /gene="nirK"
 269..281
 /note="FNR box"
 /function="transcription activator under anaerobic conditions"
 348..353
 /gene="nirK"
 364..1449
 /gene="nirK"
 /EC_number="1.7.99.3"
 /codon_start=1
 /transl_table=1
 /product="copper-containing dissimilatory nitrite reductase"
 /protein_id="CAB93142.1"
 /db_xref="GI:8249446"
 /db_xref="SPTREMBL:O9P9H9"
 /translation="MSTIPTRTRRVRVAVGTRNALAGCAAPCAKQATTEATPTQEPANAAQOTVDRIADPTAIPOIDISEKRTVSEVTTEQVALEPGVTYTMFGDQIPGMPIVRGRDITVELITINEGSMFNHIDHVRGCGAEASMPVPGQTKFRFKATPYGAFIVRPAVPLDMHISGSMGMLIVPEKGLPVDEHFFYGGHELYTGD TGEKGHDPMDEMAAEEPTVLMNGEYATLPPRHGSPSMQVEFTAVRYVTGGPNL DSSRPIGSDVDEVMQGSITAGPNRYVQTTTPVKRGSCATITLHAIVGPIKLDVHAL SRVAKATMATINREGANPDVFPEEA"
 364..456
 /gene="nirK"
 457..1446
 /gene="nirK"
 /product="copper-containing dissimilatory nitrite reductase"
 /EC_number="1.7.99.3"
 /evidence=experimental
 1570..2353
 /gene="pcn"
 1570..1575
 /gene="pcn"
 1583..2353
 /gene="pcn"
 sig_peptide
 mat_peptide
 gene
 RBS
 CDS

/codon_start=1
 /transl_table=1
 /product="proliferating cell nuclear antigen"
 /protein_id="CAB93143.1"
 /db_xref="GI:8249447"
 /db_xref="SWISS-PROT:O9P9H8"
 /translation="MDGIVRFKFMNIVASDTLQATLDSVSLVDECKIHLEDGLEIRAVDPANVGVDLRLDPAESETOGGLGVNLSTRLEDIAGNADAOQLVHLDDETRKHVISIDGLELYTALIDPDSIQEEDLPDLDSANVIGKIDRSVTAADWSDHIALGVNATDELFDYDAGSDPTDVALETRDDLILTPEDANSLPSLDVLYKMNKAIKDAEYEMLEGEFPIKMFSPFEGCGRTYMLARIQSE"
 BASE COUNT 513 a 835 c 739 g 419 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,13e-41 Length: 2506
 Score: 759.00 Matches: 161
 Percent Similarity: 55.33% Conservative: 57
 Best Local Similarity: 40.86% Mismatches: 150
 Query Match: 28.30% Indels: 26
 DB: 1 Gaps: 7
 US-10-088-045-2 (1-502) x HMA278286 (1-2506)
 QY 4 ProThrLeuIleLysThrThrLeuIleCysAlaLeu-----SerAlaLeuMetLeuSer 21
 Db 376 CCGACGGCGGACGGGAGCGGGCTTTAGAACCCCTGGCGCGGACGGCGGCACTTCGG 435
 QY 22 GlyCysSerLeuGlnAlaAspLysAlaAlaGlnProLysSerSerThrValAlaAlaAla 41
 Db 436 GGCTGT-----GCCAGCCCGCGGACGCAAAAGACCAAGCCGCA 477
 QY 42 AlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnGlnGlnGlnGlnGlnGln 61
 Db 478 GCGGAGGAGGACTCCACAGAGCCAGCTGTGAACCCGCAACACAGACCGAC----- 528
 QY 62 IleAspAlaIleValThrHisAlaProGluValProProProValAspArgAspHisPro 81
 Db 529 GTCCAGCGTATTGGCGGCGGACCCGACCGCATCCCGGACCCCATCGACCGGTGGAGCG 588
 QY 82 AlaLysValIleValLysMetGluThrValGluLysValMetArgLeuAlaAspGlyVal 101
 Db 589 AAGACCGGTGCGATGAGATGACCAACAAAGAGGATCGGAGAGATGAGACCGGCGTC 648
 QY 102 GluTyrglnPheTyrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlu 121
 Db 649 ACCTACAGTACATGACGTTGCGGACCATGATCCCGGCGCATGATTCGGGTCGCGG 708
 QY 122 GlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsnVal 141
 Db 709 GCGGACACCGTCGAACTACATCAGCAAGAGGAGGAACTGATCCGACCAACATC 768
 QY 142 AspPheHisAlaIleAlaThrGlyProGlyGlyGlnAlaGlnAlaSerPheThrAlaProGly 161
 Db 769 GACTCCACG 828
 QY 162 HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysValValValValValVal 181
 Db 829 CAGACCAAGACGTTCCGTTAAAGCTACTACCCCGCGCGGCTTATCTACCATCTGCC 888
 QY 182 ValAlaProValGlyMetHisIleAlaAsnGlyMetTyrglyLeuIleLeuValGluPro 201
 Db 889 GTCCGGAACCTCGATTCATCATCTCTCGGGGATGTTCCGGGATGATCTCGTCAGCGG 948
 QY 202 LysGlnGlyLeuProLysValAlaAspLysGlyTyrglyValMetGlnGlyAspPheTyThr 221
 Db 949 AAAAGAGGGGCTCCCGAAGTGACACGACTTCTACTTCGCGCAACAGAACTGACACT 1008
 QY 222 LysGlyLysTyrglyGlnGlnGlyLeuGlnProPheAspMetGluValAlaIleArgGlu 241
 Db 1009 ACCGCGACACCGCGGAGAGGGGACCAACGACTTCGATGAGGCGATGCGGCGCA 1068
 QY 242 AspAlaGluTyrglyValValPheAsnGlySerValGlyAlaLeuThrGlyGlu-----Asn 259

QY 311 IIEProAlaGly-----GlyAlaAlaIleThrgluPheValAspValProGlyAsp 328
 DB 1425 ATCCGGCGGGCTCGCGGGCGCGCGCTTATACCTTCAAG-----CAACCGGGCGTG 1478
 QY 329 TyrValLeuValAspIleAlaIlePheValAlaPheAsnLysGlyAlaLeuGlyIleLeu 348
 DB 1479 TACCTTACCTGACCAACCACTGATCGAGCGCTTGAAGTGGCGGGCGCGGACATC 1538
 QY 349 LysValIGluGlyGluGluAsnHisGluIle----- 358
 DB 1539 AAGGTGAGAGCGCAAGTGAACGACGATCTGATGACAGATCAAGCGCGCGCGGATC 1598
 QY 359 -----TyrSerHisLys 362
 DB 1599 CCGCGTGAAGCGCGCGCGCGCGCGCAACCGCTTGAAGACGATGATGACCC-AGG 1657
 QY 363 GlnThrAspAla-----ValTyrLeuProGluGlyAlaProGlnAla 376
 DB 1658 CAGCAAGACTCGCTGTTGGCGCGGTGCGCGCGCTGCGCGCGCGCGCGCGCG 1717
 QY 377 IleAspThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGluIleLys 396
 DB 1718 ---GACGCGCGGACGACCATCAACCGCGCGTGAATAACTAC----- 1759
 QY 397 AlaGlyLysAlaThrTyrAspSerAsnCyAlaAlaCyHisGlnProAspGlyLysGly 416
 DB 1760 -----TACAAGTACGCGCTGTGTGTGTGCGAC-----GGTCTCGCG 1795
 QY 417 ValPro 418
 DB 1796 GTGCCA 1801

RESULT 14
 AB013078 2375 bp DNA linear BCT 01-FEB-2000

LOCUS AB013078 2375 bp DNA linear BCT 01-FEB-2000
 DEFINITION AlcaI genes xylosoxidans nlr gene for Nitrite Reductase (NIR),
 complete cds.

ACCESSION AB013078.1 GI:3721763
 VERSION nlr: Nitrite Reductase.

KEYWORDS Achromobacter xylosoxidans
 Achromobacter xylosoxidans
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 Alcaligenaceae; Achromobacter.

SOURCE 1 (bases)
 Karako, K., Furusawa, H., Yamaguchi, K. and Suzuki, S.

REFERENCE Cloning and Expression of Copper Nitrite Reductase Gene from
 TITLE AlcaI genes xylosoxidans Gifu1051

JOURNAL Unpublished
 2 (bases 1 to 2375)

REFERENCE Suzuki, S. and Karako, K.
 AUTHORS Direct Submision

JOURNAL Submitted (15-APR-1998) Shimichiro Suzuki, Osaka University,
 Department of Chemistry, Graduate School of Science, Machikaneyama
 1-16, Toyonaka, Osaka 560-0043, Japan
 (E-mail: b1c@ch.wani.osaka-u.ac.jp, Tel: +81-6-850-5768,
 Fax: +81-6-850-5785)

FEATURES
 Location/Qualifiers

1..2375
 /organism="Achromobacter xylosoxidans"
 /mol_type="genomic DNA"
 /strain="Gifu1051"
 /db_xref="taxon:85698"
 590..602
 /note="fmr-box"
 822..1904
 /gene="nir"
 822..1904
 /gene="nir"
 /codon_start=1
 /transl_table=11
 /product="Nitrite Reductase (NIR)"
 /protein_id="BAA33678.1"

CDS
 /protein_id="BAA33678.1"

/db_xref="GI:3721764"
 /translation="MNALRPRLTAAALAFMAAGTAAQADAKLPTKVTTLVAPQVH
 PHEOKTKSGPKEVETIEKKNVIDKGTLLQMTENGSPPTLVVHEDYQVLT
 LVNPAITAMPINVDPEHATGALGAKLNNVDEQATLRFADRSCTGVNCAPEGV
 PMHVSGNSGSLTMDLPKGLDPOKPLIYBRYATIGFDLYIPGPRGKYDVTTLA
 ESYGTVQVMRTLTPTSHIVENGKVALTGAAALNIAKVEETVLIHSQNRDTRPHLIG
 GHGDMVETGRFANPORDLTWFIIRGSAAGALYTFKPGVAYALNHLIEAFELGA
 AGHIVKEGKMDMDLMIKIKAPAFIPR"
 sig_peptide 822..896
 /gene="nir"
 mat_peptide 897..1901
 /gene="nir"
 /product="Nitrite Reductase"

BASE COUNT 400 a 828 c 799 g 348 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,856-23 Length: 2375
 Score: 486.00 Matches: 157
 Percent Similarity: 47.22% Conservative: 55
 Best Local Similarity: 34.97% Mismatches: 176
 Query Match: 18.51% Indels: 61
 DB: 1 Gaps: 19

US-10-088-045-2 (1-502) x AB013078 (1-2375)

QY 41 AlaAlaLysThrAlaAspAlaAspAsnAlaAlaSerGlnGlnHisGlnGlyLeuPro 60
 DB 873 GCCGCGGGACCGCTGGCGCGGACGAGACGCG-----GACAGCTGCC 914
 QY 61 ValIleAspAlaIleValThrHisAlaProGluValProPro-----ProValAspArg 78
 DB 915 CATACCAAGTCAAGCTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 974
 QY 79 AspHisProAlaLysValValValLysMetGlnThrValGluLysValMetArgLeuAla 98
 DB 975 TCGGCG 1034
 QY 99 Asp---GlyValGluTyrGlnPheThrPhePheGlyGlyGlnValProGlyGlnMetIle 117
 DB 1035 GACAAAGGACAGCAGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1094
 QY 118 ArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMet 137
 DB 1095 GTGGTGCACAGAGGCGAGTACGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1154
 QY 138 ProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyGlyValGluAlaSerPhe 157
 DB 1155 CCGCAACAGTGCAGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1214
 QY 158 ThrAlaProGlyLysThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrVal 177
 DB 1215 GTCAACCTTGGGACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1274
 QY 178 TyrHisCyAlaAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMetTyr 194
 DB 1275 TACCACTCG-----GGCGCGGAGGCGATGGCGCGCGCGCGCGCGCGCGCGCG 1328
 QY 195 GlyLeuIleLeuValGluProLysGluGlyLeuPro-----Lys 207
 DB 1329 GGCACGCGTATGTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1388
 QY 208 ValAspLysGluTyrTyrValMetGlnGlyAspPheTyr-----ThrLysGly 223
 DB 1389 TACGACCGCGCTACCACTGAGGAGTTGACCTGTACATCTCCCAAGGCGCGCGCG 1448
 QY 224 LysTyrGlnGlnGlnGlyLeu-----GlnProPheAspMetGluLysAlaIleArg 240
 DB 1449 AAGTCAAGAGACTACGACAGCGTGGCGAAGCTATGGCGACAGCGTGCAGTATGGCG 1508
 QY 241 Glu---AspAlaGluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsn 259
 DB 1509 ACGTGAAGCGCGTGCAGATGCTTCAATGCAAGTCCGCGCGCGCGCGCGCGCGCG 1568

OY		260	AlaLeuValAlaValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProkln	279
Db		1569	GCGCTCACCCGCCAAGTTCGGCAGAACCCTGCCTGATC-----CATTCGACGGCCAAT	16222
OY		280	LeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGlyGly	299
Db		1623	CGCCACACCCGCCCGCATCGATCGCGGCCCATGGCCGACTGGAGTC---TGGAAGACCGGC	16779
OY		300	LysGlyGluLuan-----HisAsnIleGlnThrThrLeuIleProAlaGly-----	314
Db		1680	AAGTTCCGCCAACCCGCCGACGGCAGCACTTGAAACCTGTTCATCTCCGGCGGGTCGGCC	17399
OY		315	GlyAlaAlaIleThrGluPheLysValAspValProGlyAspTyrValLeuValAspHis	334
Db		1740	GGGCCCCGGCTCTACACTTCAAG-----CAGCTGCGGTATGCTCTACTTAACAACAC	17939
OY		335	AlaIlePheArgAlaPheAsnLysGlyValAlaLeuGlyIleLeuLysValGluGlyGlu	354
Db		1794	AACCTGATCGAGGCTTGCAACTGGGCGCGGCCGCACATCAAGAGTGAAGGGCAAGTG	18539
OY		355	AsnHisGluIleTyr-SerHisLysGlnThrAsp-----	365
Db		1854	AACAGCAGCACTGATGAAGCAGATCAAGCGCCCGCCGATTCGCGCTGACAGAGCC	19139
OY		366	-----AlaValTyrLeuProGluGlyValAProGlnAlaIleAspThrGlnGl	381
Db		1914	CGCCATGCTCCTCGCAATCAAGATTGGCACTGGCGCGCCCTGCGCTGCGCGGGCTGC	19737
OY		381	uAlaProLysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLysAlaTh	401
Db		1974	CCATGCGCGCGCGCAGCGCGCGCGCGCACTGCAC-----CCGCGCGGTGAAGACT	20244
OY		401	rTyrAspSerAsnCyshaAlaCyshIGlnProAspGlyLysGlyValProAsnAlaph	421
Db		2025	GTACCGCTCGCGCTGTGTGTGATGCGAC-----GCCAGCGCGGTGGCCACGCG--	20737
OY		421	eProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAla----SerIl	440
Db		2074	-CCCAAGCTGGCGCACAAGACGAGCGCTGGCGCCCTTCTGTGGCGCAGGCGCCGACGCGCT	21322
OY		440	evalAlaAsnGlyLeuSerGlyLys 448	
Db		2133	GCTGGCCACGGTGCTCAAAGGGCAG 2157	
RESULT_15				
RSU62291				
LOCUS			1307 bp DNA linear BCT 08-AUG-1996	
DEFINITION			Rhodobacter sphaerooides copper containing nitrite reductase (nirK)	
ACCESSION			gene, complete cds.	
VERSION			U62291	
KEYWORDS			U62291.1 GI:11480719	
SOURCE				
ORGANISM			Rhodobacter sphaerooides	
			Rhodobacter sphaerooides	
			Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;	
			Rhodobacteraceae; Rhodobacter.	
REFERENCE			1 (bases 1 to 1307)	
AUTHORS			Toques,J.E., Kwiatkowski,A.V., Shi,J. and Shapleigh,J.P.	
TITLE			Characterization and regulation of the gene encoding nitrite	
			reductase in Rhodobacter sphaerooides	
JOURNAL			Unpublished	
REFERENCE			2 (bases 1 to 1307)	
AUTHORS			Toques,J.E., Shi,J. and Shapleigh,J.P.	
TITLE			Direct Submission (26-JUN-1996) Microbiology, Cornell University, wing	
JOURNAL			Submitted (26-JUN-1996) Microbiology, Cornell University, wing	
FEATURES			Hall, Ithaca, NY 14853-8101, USA	
source			Location/Qualifiers	
			1..1307	
			/organism="Rhodobacter sphaerooides"	
			/mol_type="genomic DNA"	
			/strain="2.4.3"	
			/db_xref="taxon:1063"	

```

QY 203 GluGlyLeuPro-----LysValAspLysGluTyrValMet 215
   |||
   |||
   |||
Db 709 GACGGCTGAAGGACCAAGGAGCGCGTGCATGAGCCGCTATTATCATCGCG 768
QY 216 GluGlyAspPheTyr-----ThrLysGlyLysTyrGluGluGln 231
   |||
   |||
   |||
Db 769 GAGAGCGACCACTACATCCGGAAGACGAGACGCGACCTAC-----ATG 813
QY 232 ProPheAspMetCysValAlaIleArgGluAsp----- 242
   |||
   |||
   |||
Db 814 CGCTTCGACCCCGCGGAGCGCTACGAGACATGTCGCGGTATGACACGCTGATC 873
QY 243 AlaGluTyrValAlaPheAsnGlySerValGlyAlaLeuThrGlyLysAlaLeuLys 262
   |||
   |||
   |||
Db 874 CCACGACCATCGTCTTCAACGCGCGGTGGCGCCCTGACCGGAGGCGCGCTGAAG 933
QY 263 AlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSer 282
   |||
   |||
   |||
Db 934 GCCAAGTGGCGCAACAGTC-----CTCTTCGTC--CACAGCCAGCCCAAGCGCGACTCG 987
QY 283 SerPheIleValIleGlyGluIlePheAspLysValIlePheGluGlyLysGlyGlu 302
   |||
   |||
   |||
Db 988 CGCCGCACTCATCGCGCGGACACGCGCATCTCTGTC--TGGGAAACCGCGCAAGTTCCAC 1044
QY 303 Asn-----HisAsnIleGlnThrIleThrIleProAlaGlyValAlaIleThr 319
   |||
   |||
   |||
Db 1045 AACGCGCCGCGAGCGTACCTCGAGACTGTTCATCGCGCGGCGACGCGCGCGCGCG 1104
QY 320 GluPheLysValAlaPheValProGlyAspTyrValLeuValAspIleAlaIlePheArgAla 339
   |||
   |||
   |||
Db 1105 CTCTACAACTCTCCAGCGCGCGCTCTACGCTTACGTCACCAACCTGATCGAGGG 1164
QY 340 PheAsnLysGlyAlaLeuGlyLysValGluGlyGluLysAsnIleGluLys 359
   |||
   |||
   |||
Db 1165 GTCCACAAAGGCGCCACGCGCGCGCTCTCTCGAGGCGGAGTGGACACCACTCG--- 1221
QY 360 SerHisLysGlnThrAspAlaValTyrLeuProGluGly-AlaProGlnAlaIleAspThr 379
   |||
   |||
   |||
Db 1222 -----ATGGAACAGGTGTGTGCGCGCGTGCCTCACCG-----GCTGACCC 1263
QY 379 rGlnGluAlaProLysThrProAlaPro 388
   |||
   |||
   |||
Db 1264 GCCCGATGCGCGACCGACCGACGAGCC 1291
RESULT 16
ACNTRD 1903 bp DNA linear BCT 21-MAY-1996
DEFINITION A.cycloclastes gene for nitrite reductase.
ACCESSION 248635
VERSION 248635.1 GI:1125638
KEYWORDS nitrite reductase.
SOURCE Achromobacter cycloclastes
ORGANISM Achromobacter cycloclastes
          Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
          Alcaligenaceae; Achromobacter.
REFERENCE 1 (bases 1 to 1903)
          Chen,J.Y., Chang,W.C., Chang,T., Chang,W.C., Liu,M.Y., Payne,W.J.
          and Legall,J.
          Cloning, characterization, and expression of the nitric
          oxide-generating nitrite reductase and of the blue copper protein
          genes of Achromobacter cycloclastes
          Biochem. Biophys. Res. Commun. 219 (2), 423-428 (1996)
TITLE Cloning, characterization, and expression of the nitric
AUTHORS Chen,J.Y., Chang,W.C., Chang,T., Chang,W.C., Liu,M.Y., Payne,W.J.
JOURNAL Direct Submission
TITLE Submitted (09-MAR-1995) Wen-Chang Chang, Institute of Biological
JOURNAL Chemistry, Academia Sinica, PO Box 23-106, Taipei, TAIWAN
FEATURES
          source
          location/Qualifiers
          1..1903
          /organism="Achromobacter cycloclastes"
          /mol_type="genomic DNA"

```

```

misc_binding /db_xref="cdxon:223"
292..307
/bound_moiety="FMR protein"
CDS
475..1611
/codon_start=1
/transl_table=1
/protein_id="CA88564.1"
/protein_id="CA88564.1"
/db_xref="GI:1125639"
/db_xref="SWISS-PROT:P25006"
/translation="MTBOLMTTRRTMLAGALAGVAPLHTPAOHAAAGAAAGAP
VDISTLPRVKVDLVPVPHADQVAKTGPVVEPTMTBEKLVIDESEVHAMTF
NGSVPEPLMVHNDYVELRLINDITLLINIDFHAATGALGGALQVNGEETTL
RFKATKPGVFVHCAPEGMVPWHTSGMGAIMVLPDGLKDEKQPLTYDKIYVG
ODFYVKDEANRYKVEPTGEAYADAVKAMETLPTNIVFGAAGALGDALTAAG
ERVLVHNSQANRDRPHLIGSHGQVWATGKFRPRDIDETMLIPGSTAQAAYTFR
OPGYATVYNNHLIAFELGAAGHKVTGEMNDLMTSVKPSAM"
475..588
sig_peptide /product="nitrite reductase"
589..1608
mat_peptide /product="nitrite reductase"
BASE COUNT 369 a 601 c 604 g 329 t
ORIGIN
Alignment Scores:
Pred. No.: 1,82e-22 Length: 1903
Score: 469.50 Matches: 162
Percent Similarity: 44.11% Conservative: 55
Best Local Similarity: 32.93% Mismatches: 204
Query Match: 17.88% Indels: 72
DB: 1 Gaps: 17
US-10-088-045-2 (1-502) x ACNTRD (1-1903)
QY 19 MetLeuSerGlyCysSerAsnGlnAlaAspLysValAlaGlnProLysSerSerThrVal 38
   |||
   |||
   |||
Db 508 ATGCTGCGCGCGCTGCGCTGCGCGCGC-----CGGTTGCGCGCGCTTCCACACCGCT 561
QY 39 AspAlaAlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnLysGlnGly 58
   |||
   |||
   |||
Db 562 CAGGCGCATGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 621
QY 59 LeuProValIleAspAlaIleValThrHisAlaProGluValProProValAspArg 78
   |||
   |||
   |||
Db 622 CTGCTCGCTGCGAGAGTGCACCTC-----GTCAAGCGCGCTTGTGTGTCGAC 666
QY 79 AspHis-----ProAlaLysValValLysMetGluThrVal 91
   |||
   |||
   |||
Db 667 GCCCATGACAGGTGCGCCAGACCGGACCGCGCGCTGTGAGTTCCACCATGACGATCGAG 726
QY 92 GluLysValMetArgLeu---AlaAspGlyValGluTyrGlnPheThrPheGlyGly 110
   |||
   |||
   |||
Db 727 GAAGAAGACGTGATGATGACCGGAGGCGCATGAGTCCAGGAGATGACCTTCAAGGT 786
QY 111 GlnValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130
   |||
   |||
   |||
Db 787 TCGGTGCTGCTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 846
QY 131 AsnHisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyProGly 150
   |||
   |||
   |||
Db 847 AACCTGACCAACACAGTCTGTCACACACATTCACCGCGCAACCGGTGCGCTC 906
QY 151 GlyGlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAla 170
   |||
   |||
   |||
Db 907 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
QY 171 LeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMet----- 187
   |||
   |||
   |||
Db 967 ACCAAGCGCGCGCTTCTGTACCATTC-----GCGCGGAGGCGCATGTCCTCGT 1020
QY 188 HisIleAlaAsnGlyMetTyrGlyLeuValGluProLysGluGlyLeuPro--- 206
   |||
   |||
   |||
Db 1021 CACGTACCTCGCGCATGAACGCGCGCATCATGTGTCGCGCGCGCGCGCGCGCGCG 1080

```



```

Qy 207 -----LysValAspLysGluTyrTyrValMetGlnGlyAspPheTyr 220
Db 1081 GAGAGGGCCAGCCGCTGACCTAGACACAAAGATCTACTATGTCGGGACAGACTTCTAC 1140
Qy 221 ThrLys-----GlySerTyrGlyGlnGlyLeuGlnProhe 233
Db 1141 GTGCCGGAAGACGAGCCGGGACTACAGAGAGTACGAAACCCCGGGGAAACCTATGAA 1200
Qy 234 AspMetGluTyrValIleArgGlu---AspAlaGluTyrValIlePheAsnGlySerVal 252
Db 1201 GATGCTGTCAAGGAGGATCGGACGCTGACCCCGAACCCACATGCTTTCACAGGTCGCTC 1260
Qy 253 G1ValAlaLeuThrGlyGluAsnAlaLeuLysValGlyGluThrValArgLeuPhe 272
Db 1261 GGGCCGCTGACCGCGGACCATGCTTGTGACTCGCGCGTGGCGAG-----CGTGTCTC 1314
Qy 273 ValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheasp 292
Db 1315 GTGCTTCATTCGACGAGCCGACCGCGATACGCGCGCACCTGATCGGCGGCGATGTAC 1374
Qy 293 LysValHisPheGluGlyGlyLysGlyGluAsnHis-----AsnIleGlnThrThr 309
Db 1375 TATGTC---TGGGAGACCGGCAAGTTCGCAACCCCGCATTCGACACAGAAACCTGG 1431
Qy 310 LeuIleProAlaGlyGlyValAlaIleThrGluPheLysValAspValProGlyAspTyr 329
Db 1432 CTCATTCGCGGCGGACCGCGCGGCTCTTCAACCTTCGCGACCGCGGTGTGTAC 1491
Qy 330 ValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLys 349
Db 1492 GCCTACGTCACCAACACTATGACGCGTTCGAGCTGGTGTGCGCGCGCACCTTCAG 1551
Qy 350 ValGluGluGluGluAsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyrLeu 369
Db 1552 GTGACCGCGCAATGGAACGATGATCTG-----ATGACATCGGTGTCTCAAGCG 1599
Qy 370 ProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAla 389
Db 1600 GCGTCGATGTACCGGTCACAGCGATGATGTCGGGCGCGGTCTACAGACCCAGCC 1659
Qy 390 AsnLeuGlnGluGlnIleLysAlaGly-----LysAlaThrTyrAsp 403
Db 1660 GCGAA-CAAAAGCAGCGCGCTTGGCGGCGCTTCCCGGAGACCTTCATGCCACACCA 1718
Qy 404 SerAsnCysAlaIaLysHisGlnProAspGlyLysGlyValProAsnAlaPheProPro 423
Db 1719 GAATACCTGC-----GTTTCCGCTCT 1739
Qy 424 LeuAlaAsnSerAsp-----TyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleVal 441
Db 1740 CTCGCTTACGACGACCGGTCTGCTTTTGGCGGCAATCGCGCGTCTTC----- 1790
Qy 442 AlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerValMetPro 461
Db 1791 -----GGCGAAGGTGGGCTGTTCAGACACTCAAGAGCAGCATCGATCGTGCATG 1844
Qy 462 AlaIleAlaLeuSerAspGlnGlnIleAlaAsnVal 473
Db 1845 GCCCGCGGTACCGTGAAGATCGCGCCGACACAGATG 1880

```

```

RESULT 17
LOCUS AE007256/c 10587 bp DNA linear BCT 15-AUG-2001
DEFINITION Sinorhizobium meliloti plasmid pSyma section 62 of 121 of the
complete plasmid sequence.
ACCESSION AE007256 AE006469
VERSION AE007256.1 GI:14523793
KEYWORDS
SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE
1 (bases 1 to 10587)

```

```

AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barlcy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
TITLE Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSyma megaplasmid
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
MEDLINE 21396509
PUBMED 11481432
REFERENCE
2 (bases 1 to 10587)
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barlcy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
DIRECT SUBMISSION
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
source
1..10587
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSyma"
/complement(161..604)
/gene="azul"
/note="synonym: Sma1243"
/complement(161..604)
/gene="azul"
/function="Small molecule metabolism"
/note="glimer prediction; pseudozurxin, also called blue
copper protein, a soluble electron transfer protein
associated with periplasmic nitrite reductase;
Cu-binding/plastocyanin domain; Cu-binding/plastocyanin
domain"
/codon_start=1
/transl_table=1
/product="azul pseudozurxin (blue copper protein)"
/protein_id="AAK65336.1"
/db_xref="GI:14523794"
/transl_table="MRIAAGMAVAAVLAFTGSAPADEVRMLNKGSEGVWFEPA
FVKVNPDSVTFVPTDKSHVETIKDIPGASAFSKMETYKVTEDVGVGKCT
PHYGMGVAAVVDAPANVEKVAVALPKKARRRLDALVALAQ"
/complement(715..1362)
/gene="Sma1245"
/complement(715..1362)
/gene="Sma1245"
/function="Miscellaneous; Not classified regulator"
/note="glimer prediction; Not classified regulator"
ATG picked further upstream based on BLASTX of upstream
region, found homology to Pseudomonas"
/codon_start=1
/transl_table=1
/product="putative crp/fnr-like transcriptional regulator"
/protein_id="AAK65337.1"
/db_xref="GI:14523795"
/transl_table="MSDADIDRMILAAHATARRVPOGDAVEEGGQRTSFFLLHGRKY
TOVVEDQOIIIVRVHPGDLFGFALQKRDSDPTATAATSLASPTDIDLPQVFO
NPHLAATMOTIGORLEAHTRIREMSTOVERVAIVAVILSQAOKGEGGRIDF
PISRODIAEMTGTTLTVSRILSMBEQKGVGGRQLIICDISGLAALADGGRD"
/complement(1987..2901)
/gene="nirV"
/note="synonym: Sma1247"
/complement(1987..2901)
/gene="nirV"
/function="Small molecule metabolism"
/note="glimer prediction; global length homology to Nirv
precursor of Pseudomonas, catalyzing first step in
denitrification"
/codon_start=1
/transl_table=1

```


*
gene
CDS

/product="NirX periplasmic nitrate reductase"
/protein_id="AAK65338.1"
/db_xref="GI:14523796"
/translation="MAALSKRSKVAVPALAIPSAIILIAGLAILETGLDASPSSGA
VEEPVYTAAPDFRVRVAKEFGKNQAVADPEVTVMASALTIMKTOUTPAADVARCV
AADCAIPAEHPVPDPAMRPATIGVSFPDOAVAAMSRRGAIWVLPTBOALAFAAGS
SPFDDACAGEDBDAENPALARMDAYREARTAKARESPROBLGHGRSGSENGLSPOFN
VMEWTTCYRARITLDREGKYISDASSCGIYATGRHRAALS FVRNRGGCCAVGAPP
DNVGRLVDGSRVAPLLRTLEKGLDV"
complement(3000..4130)
/gene="nitrk"
/note="synonym: Sma1250"
complement(3000..4130)
/gene="nitrk"
/EC_number="1.7.99.3"
/function:"Small molecule metabolism; Energy metabolism,
carbon; anaerobic respiration"
/note:"glimmer prediction. Global full length homology of
NirX-like Cu-containing nitrite reductase
(disimilatory)."

/codon_start=1
/transl_table=1
/product="putative NirX Cu-nitrite reductase"
/protein_id="AAK65339.1"
/db_xref="GI:14523797"
/translation="MSEOPQMTRSMGLAGAIAGA VPLTIGIVASAHABEAVAKTAHIN
VASLPVKADVLPPEVHAHQOKEGKKYEFLITEEKIVIDEGETELHAMTFPG
SVPPGVPHODDYVELTLINDPTLOHNIDFSATGALGGALLTVNPEDTVLARF
KSKKGVPFYHCAPPGVPMHTVGSMMNGAIWVLPREGLTGKSNI YTDKYVYGEDD
FVPPNDPRFKKESVGEAYADTELVNRTLPSHIYPNGAWGLTGDSLKAQVEK
VLIIYSQNQRDTFRHLIGHGDYVATKTPRNAADVDQETPIGTGTAGAAFTTFEOP
GIYAIVNHNLIAEFELGAAPHAVTGDMNDLMTSVRPASGT"
4376..5587
/gene="Sma1252"
4376..5587
/gene="Sma1252"
/function="Miscellaneous; Hypothetical/global homology"
/note="glimmer prediction; 30% global identity to Orf 396
in the dentrificatation gene cluster of Pseudomonas
aerutina; 46% global identity to NmrS which is found
divergently transcribed from NmrR, a transcription factor
that regulates denitrification in R. sphaeroidea"

/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAK65340.1"
/db_xref="GI:14523798"
/translation="MSGKVARKRGTYVRGARTRPVLFSGRPFFLGAAWAAYVAM
TWLIAVHELEVAGSYGAAHWAHHEMLFGAPVLA GFLTAVPNMTGRLPVSGMPL
AOIFLTMILGRAALSLSDPVTHGPRAALIDGFLDLALICARVIYACRKWDKLVLOGG
LLAISLANACKHFPAVDPDHVLIARMLOISAVVALVTLIGRI LPSFTRMWLNARGST
EEPPVYNHDVTVALIAGIALGAWTLADHPTVAPVPAALHLTVRALHRMGRTMP
EMLVLYIHAYAFVPLGFAPAIGICALGVVELSWNVHYTVASALAMMLAVNTARSKH
TIPPLTAGSLTAASYAAVVLSALIRPLEAMELPEIAPTUVAGSAMILLAPLCIEYG
PLIVRRRAVQ"
complement(5643..6278)
/gene="Sma1254"
complement(5643..6278)
/gene="Sma1254"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=1
/product="hypothetical Protein"
/protein_id="AAK65341.1"
/db_xref="GI:14523799"
/translation="MSGCFPSKTNASAMQLFHGSSGRATLFRSLKSSPTTERKKR
LSIVARRCRSTRISTASSGCCOTNDIVAAGVLSAGEAPVOPHSGMNATKIYAAG
LYPFDDIKSLAPKSNWPSPSCCSIANRNAPRAPSPSTTPPCORRLPSAERVYL
LTNPILLSACVAEMCLPCRÖSSLNRITDSAPKTIASFXYRIVHSARADAPV"
5701..5970
/gene="Sma1253"
5701..5970

CDS

[illegible]

```

Oy 132 HisProAspSerLysMetProHisAsnValAspPheHisAlaAlaIleThrGlyProGlyGly 151
Db 3761 CCGGACAGACACACGCTGCAGACATATCATGCTCCACTGCGGACCGCTTGCGG 3702
Oy 152 GlyAlaGluAlaSerPheThrAlaProGlyHisIleThrSerThrPheSerPheValAlaLeu 171
Db 3701 GCGCGCGCATGACGCTGCTCATATCCGGGCGCACACCGCTGCTGCGCTTCAAGCGCGAC 3642
Oy 172 GlnProGlyLeuValTyThrHisCysAlaValAlaProValGlyMet-----His 188
Db 3641 AAGCGGGCGCTTCTGCTACCACTGC-----GACCTCCAGGAGATGTTCCGTGGCAC 3588
Oy 189 IleAlaAsnGlyMetTyrglyLeuIleLeuValGluProGlyGlyLeuPro----- 206
Db 3587 GTCACTTGGGGATGACACGGCGCATGTGTGTCGCGCGGAGGAGCATGACCGACGCG 3528
Oy 207 -----LysValAspLysGlyTyThrValMetGlnGlyAspPheTyrr--- 220
Db 3527 AAGGCAATTCATTACCTATGACAAAGGTCTACTATGTGCGGACAGACTTCTACGTG 3468
Oy 221 -----ThrLysGlyLysTyrr-----GlyGluGlnGlyLeuGln 231
Db 3467 CCGGCGGACGCAACCGCAAGTTCAAGAAATGAAAGCGTGGGGAAGCCATGCGCAT 3408
Oy 232 ProPheAspMetGluLysAlaIleArgLysAspAlaGluTyThrValValPheAsnGlySer 251
Db 3407 ACGCTCAAGTAATGACGACGCTG-----ACGCGTCCCATCATGTTCTTCAACGGCGCA 3354
Oy 252 ValGlyAlaLeuThrGlyGluAsnAlaLeuLysAlaLysValGlyGluThrValArgLeu 271
Db 3353 GTCCGGGGCTGTCACCGCGCACGCGCTGAAGGCTGCGGTGCGGAG-----AAGGTG 3300
Oy 272 PheValGlyAsnGlyLysProAsnLeuThrSerSerPheHisValIleGlyGluIlePhe 291
Db 3289 CTGATCGTCCACTCGCAGCGCCATCGGACAGCGCCCGCACCTGATCGCGCCATGCG 3240
Oy 292 AspLysValHisPheGluGlyGlyLysGlyLysAsnHis-----AsnIleGlnThr 308
Db 3229 GACTATGTC---TGGGCTACCGCGCAAGTTCCGCATATGCTCGGACGATGAGAGACC 3183
Oy 309 ThrLeuIleProAlaGlyGlyAlaAlaIleThrGluPheLysValAspValProGlyAsp 328
Db 3182 TGGTTCATACCCGCGCGCACGCGCGCTGCTTCACTTCAGTTCAGACCGCGCGCATC 3123
Oy 329 TyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeu 348
Db 3122 TATGCTTACGTCATACCATTAACCTGATCGAGGATCGAGCTTGGCGCGCGCTCATTTTC 3063
Oy 349 LysValGlyGlyGluAsnHisGluIleTySerHisLysGlnThrAspAlaValTyrr 368
Db 3062 GCGGTCACAGCGCGATGAGACATGATGATGATGATGATGATGATGATGATGATGATGATG 3004
Oy 369 LeuProGlyGlyAla-----ProGlnAlaIleAspThrGlnGluAlaProLysThr 385
Db 3003 CTGACTGAAGCGCGCTCGCGACCGCGCTCGCGACCGCGTG---GAATCGGCACAGATC 2945
Oy 386 ProAlaProAlaAsnLeuGlnGluIleLysAlaGlyLys-----AlaThrTyrrAsp 403
Db 2944 CTTGCGGCAACGATGAAGAGGATGTCGTGAGAGGCGCGCATGCGCGCATTTTCCTTCA 2885
Oy 404 SerAsnCysAla-----AlaCysHis-----GlnProAspGlyLys 415
Db 2884 GAGCAGAGAGATCGCTCCGCTGCGCTTGCATATACCTCCGCGCATTTCTGATTCGCG 2825
Oy 416 GlyValPro 418
Db 2824 GGGACTCCT 2816

```

```

RESULT 18
PSENIIRAX PSENIIRAX 1886 bp DNA linear BCT 03-MAR-1999
LOCUS Pseudomonas sp. G-179 nitrite reductase precursor (nirU) gene,
DEFINITION

```

```

ACCESSION complete cds.
VERSION M97294
KEYWORDS M97294.1 GI:151396
SOURCE Pseudomonas sp. G-179
ORGANISM Pseudomonas sp. G-179
REFERENCE Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
AUTHORS Rhizobiaceae.
TITLE 1 (bases 1 to 1886)
AUTHORS Ye,R.W., Fries,M.R., Bezborodnikov,S.G., Averill,B.A. and
TITLE A copper nitrite reductase gene and its homology to other
REFERENCE denitrifiers
JOURNAL Thesis (1992) Center for Microbial Ecology, Michigan State
AUTHORS University
TITLE 2 (sites)
REFERENCE Ye,R.W., Fries,M.R., Bezborodnikov,S.G., Averill,B.A. and
TITLE Tiedje,J.M.
JOURNAL Characterization of the structural gene encoding a
AUTHORS copper-containing nitrite reductase and homology of this gene to
REFERENCE DNA of other denitrifiers
JOURNAL Appl. Environ. Microbiol. 59 (1), 250-254 (1993)
PUBMED 8439151
JOURNAL 3 (bases 1 to 1886),
PUBMED 8439151
REFERENCE Ye,R.W.
AUTHORS Direct Submission
TITLE Submitted (26-APR-1993) Life Sciences, Dupont, Wilmington, DE
JOURNAL 19880, USA
FEATURES
source Location/Qualifiers
1..1886
/organism="Pseudomonas sp. G-179"
/mol_type="genomic DNA"
/strain="sp"
/isolate="G-179"
/db_xref="taxon:71238"
/note="similar to Rhizobium"
289..1807
/gene="nirU"
289..303
/gene="nirU"
/standard_name="PNR box"
/note="putative"
/bound_moiety="FNR protein"
/function="activates transcription"
580..595
/gene="nirU"
/standard_name="NtrA box (sigma-54 binding site)"
/function="facilitates recognition of RNA polymerase to the promoter"
623..639
/gene="nirU"
/standard_name="NtrA box (sigma-54 binding site)"
/function="facilitates recognition of RNA polymerase to the promoter"
654..660
RBS the promoter"
654..660
CDS /gene="nirU"
668..1807
/gene="nirU"
/codon_start=1
/transl_table=11
/product="nitrite reductase precursor"
/protein_id="AAC79132.1"
/db_xref="GI:151397"
/translation="MSDFRITRSMAGAAVAGALAPVTVTSVAHABGGGKITNSAAT
AANTALERVKVELVKPPFVAHQKAGEBPVAFKPTIQEKI VDDDKTEVHAMT
FDGSGVPGPMIVHQDDYVELTVNPDNTELDHNSATGALGGALTVNPDYAV
LRFATKAGVYVHCADPAGWPMVTSGMNGALIVLPDDGDKHKGHELVYDKVYVG
GDVYVYKDENGKFKYKESAGVAVDYLKAMKTLPPHVVENGAVGALTGNAIOAKV
GDRVLLHSQANPDTRPHLITGCHGDVYMAATGKFPANPELDETWFIPGGAAGAAVYTF
QQQPIYAVVNNLLIEAFELGAGAFKVTGDMNDLMTAVVSPISG"
668..763
/gene="nirU"
sig_peptide

```

mat_peptide

764. .1804

```

/product="nitrite reductase"
/function="nitrite reduction"
/name="Cn vld_nitrite"

```

BASE COUNT	439 a	501 c	554 g	392 t
ORIGIN				

Alignment Scores:

Pred. No.:	2,286-21	Length:	1884
Score:	453.00	Matches:	118
Percent Similarity:	51.58%	Conservative:	45
Best Local Similarity:	37.34%	Mismatches:	119
Query Match:	17.25%	Indels:	34
DB:	1	Gaps:	10

US-10-088-045-2 (1-502) x PSENIRAX (1-1886)

QY	72	ValProPProPProValAspArgAspHis-----ProLalysVal	84
Db	842	GTCAAGCCACCTTTGTGACAGCCCCACACCAGAAAGCTGAGAGCGGAGCCCAAGTTTCTC	901
QY	85	ValValIlyMetGluThrValGluIlyValMetArgLeuAlaAsp---GlyValGluIyr	103
Db	902	GAGTTCAAGATGACCATCCATCCAGGAAAAAGAAATCGTTGTGCAGCAGCAAGGGAGCCAGAGTC	961
QY	104	GlnPheTrrPThrPheGlyGlyGluValProGlyGlnMetIleArgValArgGluGluAsp	123
Db	962	CATGGCATGACCTTGACCGAGATCTGTGGCGGGGCGAGATATATCGTTTACCAAGATGAT	1022
QY	124	ThrIleGluValGlnPheSerAsnHisProAspSerIlyMetProHisAsnValAspHe	143
Db	1022	TATGCGAATGACCTCGTCATATCCCGATCAAAAGATTCAGACAGCAATATCGACTTC	1081
QY	144	HisAlaAlaThrGlyProGlyGlyGlyValGluAlaSerPheThrValAlaProGlyHisThr	163
Db	1082	CATTGCGAGCGGGTGGCCCTGGTGTGTGAGCGCTGACCGTGCATATCCCGTGTACAG	1144
QY	164	SerThrPheSerPheIlyAlaLeuGlnProGlyLeuIyrValIyrHisCysValAlaVala	183
Db	1142	CGCGTCTCGCTTCAAGGCCACCAAGCGGGTGTTTGTCTATCATCTG-----GCC	1199
QY	184	ProValGlyMet-----HisIleAlaAsnGlyMetIyrGlyLeuIleLeuValGlu	200
Db	1196	CCCCCGGCGATGCTGCCGTGCATGTCACTTCGGCGCATGAATGGTGCATATGATGCTCCG	1255
QY	201	ProIlyAsnGluIlyLeuProIlys-----ValAspIlyGluIyrIyr	213
Db	1256	CCGGCTGACGGCTCTCAAGATCACAAGGCCACGAGCTTTTACGACAAAGGTCTACTAT	1311
QY	214	ValMetGlnIlyAspPheIyrThrIlys-----GlyIyrIyrGly	226
Db	1316	GTCCGGCGACGAGACTTTTATGTCCCGAAGATGAGAACCGCAATTCAGAAATACGAA	1377
QY	227	GluGlnGlyLeuGlnProPheAspMetGluysAlaIleArgGlu--AspAlaGluIyr	245
Db	1376	AGCCCGCCGACGGCTATCCGATGTGCTGCGAAGCATGAAACATGACACTGACCCCGAGCAT	1433
QY	246	ValValPheAsnGlySerValGlyAlaLeuThrGlyValAsnAlaLeuValAlaIysVal	265
Db	1436	GTCTGTTTCAACGGGGCGGCTCGGCCCTTGACGGCGCATTAAGCTTGAAGCCAAAGTG	1495
QY	266	GlyIlyuThrValArgLeuPheValGlyAlaAsnGlyIyrProAsnLeuThrSerSerPheHis	285
Db	1496	GCGCAT-----CGTGTCTGATTTCTTATTCGCAAGCAACAGGATATACGGCCCGCAC	1544
QY	286	ValIleGlyGlyIlePheAspIlyValHisPheGluGlyGlyIlyGlyGluAsnHis---	304
Db	1550	CTGATCGGGGGGACATGGCATTAATGTC--TGCGCTACCGGCAATTCGCCAACCCGCGG	1606
QY	305	-----AsnIleGlnThrThrLeuIleProAlaGlyGlyValAlaIleThrGluIlyHis	322
Db	1607	GAACTCGATCAGGAAACTGTGTTCTATTTCCGAGAGTCTGCTCCGGGGCGGCTTACTACACG	1666

Oy 323 ValnspValProglYasPTyValleuValnshAspiSaIaiEneArxAlApheslnys 342
 Db 1667 TTCACGACGCCCGGTATTCCTGATTGAACAACAATTCGATCGAGCGCTTCAACTC 1722
 Oy 343 GYAIAaleugiyileuLyVaIGluGIyGlubshniSgiulle 358
 Db 1727 GGCCGGCGGCACCTTCAAGTGACGGCGCATCGAAACGACGACTG 1774
 RESULT 19
 AF040987
 LOCUS
 DEFINITION Pseudomonas sp. G-179 nitrite reductase (nirV), NirV precursor
 ACCESSION AF040987
 VERSION
 KEYWORDS
 SOURCE Pseudomonas sp. G-179
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae.
 REFERENCE
 AUTHORS 1 (bases 1 to 4029)
 TITLE Ye,R.W. and Laura,B.,
 Characterization of the DNA region responsible for nitrite
 reduction in Pseudomonas sp. G-179 containing a Cu-type nitrite
 reductase
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4029)
 AUTHORS Ye,R.W. and Laura,B.
 TITLE Direct Submission
 JOURNAL Submitted (01-JAN-1998) Environmental Biotechnology, E328/148B,
 Dupont Central Research & Development, Route 141 and Henry Road,
 Experimental Station, Wilmington, DE 19880-0328, USA
 FEATURES
 source location/Qualifiers
 1..4029
 /organism="Pseudomonas sp. G-179"
 /mol_type="genomic DNA"
 /strain="G-179"
 /db_xref="taxon:71238"
 582..1721
 /gene="nirV"
 /note="Cu-type nitrite reductase structural gene"
 582..1721
 /gene="nirV"
 /codon_start=1
 /transl_table=1
 /product="nitrite reductase"
 /protein_id="AAB86769.1"
 /db_xref="GI:2773328"
 /translation="MSBOFLRTRSMLAGAVAGALAPVTVSAHAEQGIKINSAAI
 AAIATLTERKVELVKRPFPFAHTOKRAGEPKVEFPGTIQEKIVDDKTGEVAMTIT
 FDSVPSPMIIHODDYVELTLVNPDGNLOHNIDPSATGALGLGALTVPDPA
 LRKRTAGVAVFHCAAPGMVPMHVNTSGNAILMV.PRDGLKHKGHEIAYDKRYVVC
 EQPPYVDXENGSKFKKYESNGEAYPDVLEAKMTITPHVVFNCAVGALTGDNAIQARK
 GDVLILHSOANDIRPHLIGHGADYVAWGKRANPELDQETWFIFGAAGAAYTFTE
 QOGPIAYVANHNIEAFELGAAGHFVTGWMDMLTRAVVSPTS"
 1803..2699
 /gene="nirV"
 1803..2699
 /note="similar to cystathionine gamma-synthase"
 /gene="nirV"
 /codon_start=1
 /transl_table=1
 /product="NirV precursor"
 /protein_id="AAB86770.1"
 /db_xref="GI:2773329"
 /db_xref="GI:2773329"
 /translation="WTFRDAFFRNVLSTAFPALVVGLTGAIGFOSSGLPTRERG
 PVATISIPAGLLEYBSGEFFPNGLVDPMPKRIHOSLEIMKFVSGESYDCVAB
 GVCAARDLSGSRADMPATGSFDDAAAYVLSNRNSGFWLRPLKEIELFAASGYF
 DMLGIDPSIKRMALRWLYERETARPARISYPQVGFGENHEGLADFGNIWEW
 TSCICLRVTLDRYSKPSISEASGCIYIATGKHAPLSFTIREPKGCSGVAPPDNVG
 FRVVRPRWTPLLPALISGRSM"
 2840..3511

```

/ gene="nmr"
/ gene="nmr"
/ note="similar to Nmr from Rhodobacter sphaeroides and
other FNR-like regulators"
/ codon_start=1
/ transi_cable=11
/ product="Nmr"
/ protein_id="AB96771.1"
/ db_xref="GI:2773330"
/ translations="MPDELDRLDLSARFAGDPAVEGGPADSYLLHGRUKV
TQVBERGQOIIYRVVHPDLFGFARLQKRDYATVAVESTLAMPETLMSFVH
NPLAVTAQITGERLEBAHVIREMSIEVERVAHVIRLARKKQKQKDIRDF
LISRODIAMTGTTHATVSRIISAWOLGFGVGRKLPVQISGLTASGEFRRO
GTASQ"
BASE COUNT      864 a 1110 c 1203 g      852 t
ORIGIN
Alignment Scores:
Pred. No.:      5.4e-21      Length:      4029
Score:          453.00      Matches:      143
Percent Similarity: 45.93%      Conservative: 60
Best Local Similarity: 32.35%      Mismatches: 176
Query Match:     17.25%      Indels:      64
DB:              1          Gaps:      15
US-10-088-045-2 (1-502) x AF040987 (1-4029)
QY 72 ValProProProValAspArgAspHis-----ProAlaIysVal 84
DB 756 GTCACGCCACCCCTTGTGCACGCCACCCAGAGGCTGAGGCGAGCCCAAGGTTGTC 815
QY 85 ValValIysMetGluThrValGluIysValMetArgLeuAlaAsp---GlyValGluTyr 103
DB 816 GAGTTCAAGATGACCATCCAGAAAGAGATGCTGTCACACACAGGAGGACCGAGTC 875
QY 104 GlnPheTyrThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGluGlyAsp 123
DB 876 CATGCAGATGACGTTTCAGCGATGTCGCGGCGCGCATGATGATCTGCACAGATGAT 925
QY 124 ThrIleGluValGlnPheSerAsnHisProAspSerIysMetProHisAsnValAspPhe 143
DB 936 TATGTGAACTGACCTCTGTCATCCGATVCAAAACAAATTGACAGACATATCGACTTC 995
QY 144 HisAlaAlaThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAlaProGlyHisIsthr 163
DB 996 CATTCGCGACAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1055
QY 164 SerThrPheSerPheIysAlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAla 183
DB 1056 GCGGTGCTGCGTTCAAGGCCACCAAGCGCGGTGTTTGTCTATCATCTGT-----GCC 1109
QY 184 ProValGlyMet-----HisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlu 200
DB 1110 CCCGCGGCAATGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1169
QY 201 ProIysGluGlyLeuProIys-----ValAspIysGluTyrTyr 213
DB 1170 CCGCGGAGCGGTTCAGAGATCCAGAGGCCAGAGCTGTTTACGCAAGAGGCTTACTAT 1229
QY 214 ValMetGlnGlyAspPheTyrThrIys-----GlyIysTyrGly 226
DB 1230 GTCGCGGAGAGGACTTTTATGTCGCGAGAGATGAGAAGCGCAAGTTCAAGAAATGCGAA 1289
QY 227 GluGlnGlyLeuGlnProPheAspMetGluValAlaIleArgGlu---AspAlaGluTyr 245
DB 1290 AGCGCGCGAGGCGCTATCGAGTGTGCGAGGCGCTGAAGCACTGACGCCGAGCGCAT 1349
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuIysAlaVal 265
DB 1350 GTCTGTTCAACGGGGCGGTGCGCGCTTCAAGCGGCGCATTAACGCTTTCAGGCGCAAGGTG 1409
QY 266 GlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHis 285

```

```

DB 1410 GCGGAT-----CGTGTCTGATTTCTTCATTCGCAAGCCAAACGAGATACGCCCGCAC 1463
QY 286 ValIleGlyGluIlePheAspIysValHisPheGlyGlyIysGlyAsnHis--- 304
DB 1464 CTGATCGGGGGGATGCGCATTTATGTC---TGGGCTACCGGCAAGTTCCGCAACCGCGCG 1520
QY 305 -----AsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGluPheIys 322
DB 1521 GAACGTGATGAGAAACGTGTTTCATTTCCCGAGAGTGTGCGCGGGGCGCTTACTACAG 1580
QY 323 ValAspValProGlyAspTyrValIleValAspHisAlaIlePheArgAlaPheAsnIys 342
DB 1581 TTCACAGACCGCGGTATCTATGCGTATGAAACACAACTGTATCGAGGGGTTCGAACTC 1640
QY 343 GlyAlaLeuGlyIleLeuIysValGluGlyGluAsnHisGluIleTyrSerHisIys 362
DB 1641 GCGCGCGCGCGGACCTTCAGAGTACGCGGCGGACTGGAACGACGACTGATGACA----- 1694
QY 363 GlnThrAspAlaValTyrLeuPro-----GluGlyAlaProGln 379
DB 1695 -----GCGGTGTTTCGCGGACCTGCGGTGACCGGTGACCGGTGCGCGCGGATCA 1745
QY 376 AlaIleAspThrGlnGluAlaProIysThrProAlaProAlaAsnLeuGlnGlnIle 395
DB 1746 GCGCGCGCGGCAACATCTTATGCGCGCGGAGGCGCGGAGATCGAGATCCCATG 1805
QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnCysAla----- 407
DB 1806 ACCTTTCGCGACGCTCTTCTT-TGCAATGTTGTTTTCGATGCTTTCCTTCGCTTCCG 1864
QY 408 AlaCysHisGlnProAspGlyIysGlyValProAsnAlaPheProProLeuAlaAsnSer 427
DB 1865 GCTTGTGTCGCGCTACCGCGCG-----ATTCGCTTTCAGTCCGAGCTTTTGCC 1915
QY 428 AspTyrLeu-----AsnAlaAspHis-AlaArgAlaAlaSerIleValAl 442
DB 1916 GACGCGCGAAAGGCGCGTCCGCTTCGCGGACATCATGATCCCGAGCTCTTTCGGA 1975
QY 442 AsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerValMetProAl 462
DB 1976 ATATCGAAGAAAGTGAGTCTTCCGCAAGGCTTCACCGTGCAGCGCGGATGACCA 2035
QY 462 attt 463
DB 2036 AATT 2039
RESULT 20
AB008247 10484 bp DNA linear BCT 18-DEC-2001
LOCUS Agrobacterium tumefaciens str. C58 linear chromosome, section 51 of
DEFINITION 187 of the complete sequence.
ACCESSION AB008247 AB007870
VERSION AE008247.1 GI:15158849
KEYWORDS
SOURCE Agrobacterium tumefaciens str. C58 (Cereon)
ORGANISM Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
REFERENCE 1 (bases 1 to 10484)
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10484)
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Direct Submision
JOURNAL Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
Street, Cambridge, MA 02139, USA
COMMENT Approximately 800 bp of telomeric sequence missing from the left
end of the chromosome and 200 bp missing from the right end.
FEATURES
Location/Qualifiers

```

source
1. 10484
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/mol_type="genomic DNA"
/strain="C58"
/db_xref="taxon:181661"
67. 1965
/gene="AGR_L_961"
67. 1965
/gene="AGR_L_961"
/note="(AF083948) TrkA-like protein"
/codon_start=1
/transl_table=11
/product="AGR_L_961p"
/protein_id="AK89054.1"
/db_xref="GI:15158850"
/translation="MLDFLEETVYGRAMHRLIGETGTWLRFPDHAVALLEDMOAVLTI
CFRGCEHVAOIAAPAPARTSTHRLRLQKRGICGEKLAQGRPHATLMLEPVLDLPI
DRLRNRLIYLMALMALMLLEPVAAADPLNLDLRLKRLKRLAEIYTRAPFKTKRY
RRLCAAMLAIVQTRPLPQAEQVQEVFLFMAAAGIEGVAPEPASPEKAPRPLML
PVLPMDTLRLRETSARQEDQPAAPQEAARHIAVREKAKENRQTERSPFLINR
FEKILAAEMVNVDRPGDSDADGSADELDLTLGKGRPSARFPFDLDLPEAL
NHTALTAELTYPEMDYRERYTLADHCRVTSIPAKDAETAPEDPDTKMLIRVRROE
VLRPRHLLRAOLDGNDLDLAVRARSDIAAGCGSRHIVASRPOANDLAATLVD
VSLTDAMFDRRLVDKEKLLVLAHGLSCGTHSITLTSTRSRVRYETKDFD
EPMGHTLERRIAALKPGFYTRIGPAIRHSAKLERPERRLLLLDGKNDVDYHE
GRFAIEDSRRAVSEARRSGVFGVTDSKQSYIPAMFGNGYAIIVRIKLPALP
AIRSLAG"
1965. 2414
/gene="AGR_L_963"
1965. 2414
/gene="AGR_L_963"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_963p"
/protein_id="AK89055.1"
/db_xref="GI:15158851"
/translation="MLVTNLTJNRFVCAATMSPNGTADRFQSGAPRHRTSFHPP
WSQMSKQATNIDVRIIPRERHPRIFGALGALTEGSLQITSDHDPRLPLVQLAT
NPGQFGEWYLEQGPVWRDLIGRLBEQABESSEDMSCGCCGSEH"
2493. 3026
/gene="AGR_L_966"
2493. 3026
/gene="AGR_L_966"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_966p"
/protein_id="AK89056.1"
/db_xref="GI:15158852"
/translation="MTEWIRELDVRLPLQSGGPPPAIMEAVAGLAPCEALRLYATFR
PVLPKMAERGFDOHREIGGDMEVLFTQSIETEOISVPALAGNAEERPOIWWPD
AIVLDSEETAGCOTSRKSLDLRLPGAVIFVLMPOEPTLYPELLARGHONAGNPD
EARTAFRIIRAGAGAH"
complement(3110. 4384)
/gene="AGR_L_968"
complement(3110. 4384)
/gene="AGR_L_968"
/note="conserved hypothetical protein PA2662 (imported) -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_968p"
/protein_id="AK89057.1"
/db_xref="GI:15158853"
/translation="WRLDICODQARSTTMSSGVTVPLGSRPKGIPRLGARTGPVI
FSYGFRRPFLGALMAIVAMVLAISGFLDIGDGLVDPGVVWAAVADGFLPALS
FLTRAVNMTGRLEPVSGKPLVWLFWLWCAGRLLLDDPOVGVVAAVADGFLPALS
ICAREVIAGRKMDLKVGLALASVANIYFHVAAIGDHSOMATRLASVATYLVII
VGRIVPSFTNNMLNRFGRTPFPVNGFPAALITGLVALAVMTIEPESMLAVSTLL
LAAEMHVRLLRMGWTTRPEQVIVLHVAVAFIPVGFIAIALADVMQDRSYLHF
TVGVIGCMALAVMTRASIGHTGRKLASRLTIAYVSLIACALLPAAEFLPGVMHL
YGCALIMIVGFLFVEYGPILMRERPLKA"

gene
4354. 5721
/gene="AGR_L_970"
4354. 5721
/gene="AGR_L_970"
/note="COPPER-CONTAINING NITRITE REDUCTASE PRECURSOR
(CU-NIR)"
/codon_start=1
/transl_table=11
/product="AGR_L_970p"
/protein_id="AK89058.1"
/db_xref="GI:15158854"
/translation="MLGLDICLDSPEGFPANRQGGATLPPOPSAACONVAGRRKS
LFHLKLYFRSSATNAVATERNMLCEKRDNRMMETFRHITRNILAGAAFGALAPMI
PSAAGAEEKKAAPLTSALPAKVDLVKPPVHAHTQAKGEPKPIVEFTLTI
KEOMILDDKGTETVHMTFNGSVGPMVHODDYVELTLINPTNELQHNIDFSAT
GALGGGLTIVNGEXAILRPKATKGVFVYHCAPCPMVMMHTSGMNGAIMVLPREG
LTGCHGKELYVDKYVYGGEDPITPRDENRPFKYESAGCMADTLEVMKLTSHYV
FNGVGLVTEHLDQAAVSGKVLIVHSQANRDTRPHLIGHGDTVMAKFRNPDDID
QETWFIIPGTAGAAFYTFEQPIYAVVNNLIEAFELGAAAHFKVTGEMNNITLMQAVL
APSSI"
5800. 6747
/gene="AGR_L_972"
5800. 6747
/gene="AGR_L_972"
/note="(AF040987) Nirv precursor"
/codon_start=1
/transl_table=11
/product="AGR_L_972p"
/protein_id="AK89059.1"
/db_xref="GI:15158855"
/translation="MASPRRQVLMGIKGNSHPEPNVSLPLSLATVLTATAIFQSG
IIRSGADTRAMTAPQVTVYVAGFTYRAGEYFRDQVAVDGPVLDLTKRONLIMKY
QVSSANYARCVAAGCCQPEGFAAPVGGSIPTAGISODDAHAYAKLSRTGAVML
PSDAELAFAGSLPDDETLGVADSENPAIRMLADYRQTSRAAGNPVQSGSGE
NEYGLADPGCNIMETWTRSCSHRTTLKGDSGVDTTAAACGIPVAGSKRAAMSAFVKP
KSGCAVGPDDMLGFRLVKDTWYAPLVEKLPAALIS"
6891. 7598
/gene="AGR_L_974"
6891. 7598
/gene="AGR_L_974"
/note="(AB031072) FNR-like protein
partial sequence"
/codon_start=1
/transl_table=11
/product="AGR_L_974p"
/protein_id="AK89060.1"
/db_xref="GI:15158856"
/translation="MVKIDRSVIRSLAFARADELDGLTHANSRLVPPGEAITE
QGGSAHFLLIHGRKLVTVTHDGOIIVRMVPGDLFGFARALORDSDYPTATTA
ESVILSWQTDLPQFVQENPHLAVSAMQITGRLLEAHTRIREMSTOEVERVAHAVL
RLSQOGRKELDIRIDFPIRSDIDAMETGTLHTVSRILSSWSKGLVCGGRKLLV
CNLPGLAQLAEGSTQV"
complement(7669. 8088)
/gene="AGR_L_976"
complement(7669. 8088)
/gene="AGR_L_976"
/note="CYTOCHROME C2"
/codon_start=1
/transl_table=11
/product="AGR_L_976p"
/protein_id="AK89061.1"
/db_xref="GI:15158857"
/translation="WRQDKDQITRNISIIAAMLAATVPALSEGVAKGEAVFKRS
ACHAIGCAKNRVPOLNGIIGRAAGVPDYTYSAMKKAGDEGLVWTPPELNDPLSA
PKKIKGNKALAGISKPELDNLIAVIESTASKRAE"
complement(8134. 8697)
/gene="AGR_L_978"
complement(8134. 8697)
/gene="AGR_L_978"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_978p"

/protein_id="AAK89062.1"
 /db_xref="GI:15158858"
 /translation="MNNRRLKLIATATGLPLIGADVLTAAENNAKPAAGGVHSP
 EIRFLDEAVATIIPTSTPGAKDAVGAFAVAADYQDEBQALPLSAIPEIKESQ
 ADHKFGLDLEOKALISALDKOABQETPAKPAHFTLVKQTLTLLGFTSTKATGE
 VLVDYDIKPGFEDRVPYKKGTPAWGTT"
 complement (8694..10379)
 /gene="AGR L 980"
 /complement (8694..10379)
 /gene="AGR L 980"
 /note="probable oxidoreductase chain Cj0415 (imported)
 Campylobacter jejuni (strain NCTC 11168)"
 /codon_start=1
 /transl_table=11
 /product="AGR L 980p"
 /protein_id="AAK89063.1"
 /db_xref="GI:15158859"
 /translation="MADNHYDAIVVSGISGMAAELOKGLKTLMEERGNIEVT
 DYONADKEADYPHNRATQEMKAPVLSRDYLBLEATGMADQETPYVEKRPD
 WRGYVVGSRSLIMGQTYMSQDEPAANKDGIADWPIRYEDVSPWYVERFAGI
 SGRBGLDILPDEFLPIPLNFVEDVASRLKAKFKGTMLNSCANITOLPOE
 RFPCFRNKRCLGCPFGYVSTOASTLPAAVATGNLTLPFSIYKELVDKKKRG
 VRLDAETLVYETADIIETANSTLNSPVMNSADTWEGGLSSSGELGNMWDH
 HRRMGATGVDFEDEFYFKRRPAGYIIPFRNTGDKKTYLGFQISASRMR
 EIAELNIGADYKDTLEPGWTIGMTAFGMLPYHNDNRVLDHDKDKLPGVLSNV
 EKONLEDMREDMVDVMEFAVGIKNYVSRGTYPAGMGIHEMGTARGLPKTSV
 LNNQNVMDAPNVFTDGAICTASCNPSLTLYMALTAARAEFAVERKKGNL"

BASE COUNT 2080 a 3290 c 3165 g 1949 t

Alignment Scores:
 Pred. No.: 1,61e-20 Length: 10484
 Score: 453.00 Matches: 148
 Percent Similarity: 43.43% Conservative: 67
 Best Local Similarity: 29.90% Mismatches: 168
 Query Match: 17.25% Indels: 112
 DB: 1 Gaps: 17

US-10-088-045-2 (1-502) x AE008247 (1-10484)

QY 21 SerGlyCysSerAsnGlnAlaAspLysAlaGlnProLysSerThrValAspAla 40
 DB 4669 GCGGGCGGCGCAAGAAAGAGCCGCAAGGCGGAGATTGCA 4728
 QY 41 AlaAlaLysThrAlaAsnAlaAsp-----AsnAlaAlaSerGln 53
 DB 4729 GCGCTCCCTCGCGCAAGGTTGATCTCGTCAAGCCCTTCTGCGCAACCCAG 4788
 QY 54 GlnHisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaProGluVal 73
 DB 4789 AAAGCCGAAGCGGCGCAAGATCGTCAAGCTCACGCTCACG----- 4830
 QY 74 ProProValAspArgAspHisProAlaLysValValLysMetGluThrValGluLys 93
 DB 4831 -----ATCAAG-----GAAACG 4842
 QY 94 ValMetArgLeuAlaAsp---GlyValGluTyrGlnPheThrPheGlyGlyGlnVal 112
 DB 4843 AAGATATCTCTCAACGCAAGGCGCAAGGCGGATTCAGCCATTCACAGGCTGGGT 4902
 QY 113 ProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHis 132
 DB 4903 CCCGGCCCTGTATGTGTGGTGCATCAGAGACTAGCTGAGCCCATCAACACCC 4962
 QY 133 ProAspSerLysMetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyGly 152
 DB 4963 GATACCAAGAGCTGACGCAACATCATGATTCACCTCGGCGCACCGCGGCTCGGCGGC 5022
 QY 153 AlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGln 172
 DB 5023 GCGGGGCTCAACATCTGCAACCCCGGGAAGAGCGGATCTCTGCTTCAAGGCCACCAAG 5082
 QY 173 ProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMet-----HisIle 189

DB 5083 GCGGGGCTCTTCCTTATATCACTGC-----GCACCTCCCGGCGATGCGCGTGCATGTC 5136
 QY 190 AlaAsnGlyMetCysGlyLeuIleLeuValGluProLysGluGlyLeuProLys----- 207
 DB 5137 ACCTCGGCGATGATGCGCGGATGATGATGCTGCGCGCGGAGGCTTCAAGCAGCGGCAC 5196
 QY 208 -----ValAspLysGluTyrTyrValMetGlnGlyAspPheTyrThrLys 222
 DB 5197 GCGAAGAGCTTGTATTATACAAAGCTATTATGTTGGGAGAGAGACTTCTTACATCCG 5256
 QY 223 -----GlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMet 235
 DB 5257 CGTGACGAGACGCGCAATTTCAGAAATATGAAGCCGCGGTATGCCATGCGCGACACG 5316
 QY 236 GlnLysAlaIleLeuArgLys---AspAlaGluTyrValValPheAsnGlySerValGlyAla 254
 DB 5317 CTGGAAATGATGCGCAAGCTGACGCGCAAGCCACATGCTTCAATGCGCGATCGCTGCC 5376
 QY 255 LeuThrGlyValAsnAlaLeuLysValLysValGlyValThrValArgLeuPheValGly 274
 DB 5377 CTGACAGGGGAACATCCCTTCAAGCGCGCTGCGTGGAA-----AAGTGTGATGCTGT 5430
 QY 275 AsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLysVal 294
 DB 5431 CATTCGACGAGCGCAACCGCGACACCCCGCGCATCTATCGCGCGCACGCGCATACGTC 5490
 QY 295 HisPheGluGlyGlyLysGlyLysAlaHis-----AsnIleGlnThrThrLeuIle 311
 DB 5491 ---TGGGCGACGCGCAAAATTCGCAACCCCGCGATCTCGACGAGAAACCTGTTTCATC 5547
 QY 312 ProAlaGlyValAlaAlaIleThrGluPheLysValAspValProGlyLysPheValLeu 331
 DB 5548 CCGGGGCGACGCGCGCGCGCGCTTCTTACCTTGGAGAGCGCGGCACTTACGCTAC 5607
 QY 332 ValAspHisAlaIlePheAlaArgAlaPheAsnLysGlyValAlaGluGlyIleLeuLysValGlu 351
 DB 5608 GTCACACCAACACTCATCCAGAGCTTCGAACTGGGCGCGTGCAGCCCATTCAGAGTCACG 5667
 QY 352 GlnLysGlnAsnHisGluIleLysSerHisLysGlnThrAspAlaValTyrLeuProGlu 371
 DB 5668 GGTGAAGTGAACAACACCCCTG-----ATCGAGGCGGTGCTC----- 5703
 QY 372 GlyAlaProGlnAlaIleAspThrGln---GlnAlaProLysThrProAlaProAlaAs 390
 DB 5704 ---GCCCATTCAGAGCTTGAAGCGCAACGAGGAGCGCGCGCGCA----- 5746
 QY 390 IleGlnGlnGluIleLysAlaGlyLysAlaThrTyrAspSerAsnCysAlaIleCysHis 410
 DB 5747 -----GCGCGGCGCG 5757
 QY 410 SerLysProAspGlyLysGlyValProAsnAlaPheProLeuAlaAsnSerAspTyrLe 430
 DB 5758 CTCCTCCCTTCGCTCTCTCGCGCTTGAAGCGCGCAACTTTCGCTTCCAC----- 5812
 QY 430 AsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerCysLysIleThr 450
 DB 5812 ----- 5812
 QY 450 ValAlaGlnLysAsnGlnTyrGluSerValMetProAlaIleAlaLeuSerAspGln-Gln 470
 DB 5813 ---GACGAGGAGAGGTTTCATGCGGATTCGCGGCAATATGCAATTCCTCAATGCGCTC 5868
 QY 470 LeuAlaAsnValIleThrTyrThrLeuAsnSerPheGlyAsn 483
 DB 5869 TCGCTTCTTTGCGCGCTCTCGCTTGCACAGTGCCTGCGCAAC 5909

RESULT 21
 AE009367/c 10753 bp DNA linear BCT 20-DEC-2001
 LOCUS Agrobacterium tumefaciens str. C58 linear chromosome, section 137
 DEFINITION of 187 of the complete sequence.
 ACCESSION AE009367 AE008689


```

Db      1095 CTCCCTTCGCTCTCTGCGCGTGAAGCAGCGCCCAACTTGGCGTCCAC----- 1041
Qy      430 uAaenlaaaphislaaargalaaaserilaValaalaennglyseusergllylsileth 450
Db      1041 ----- 1041
Qy      450 rvaAaennglyAaennglyrGluSerValMeProAlaIleAalaenSerAepGln-Gln 470
Db      1040 ----GAGCAGCAGCAGGTTCTCATGGGCAATCCCGGCATACCAATTCCTCCAAATGCCGTC 985
Qy      470 leAaenValIlethryrThreleuAenSerPheglyAaen 483
Db      984 TCGCTCTTTTGGCGCTCTCGCTTGCACAGTGTGGCAAC 944

RESULT 22
AF083948/c 20761 bp DNA linear BCT 28-APR-1999
LOCUS      AF083948
DEFINITION Pseudomonas sp. G-179 FNR-type regulatory protein (ntrR), NtrV
            (ntrV), Cu-type dissimilatory nitrite reductase (nirV), NapB
            (napB), ferredoxin-like protein (napF), NapD (napD), periplasmic
            nitrate reductase large subunit precursor (napA), periplasmic
            nitrate reductase small subunit precursor (napB), cytochrome c
            binding protein (napC), anaerobic coproporphyrinogen III oxidase
            (hemN), trkA-like protein (ntrD), putative chaperone (ntrQ), nitric
            oxide reductase large subunit precursor (norC), NorF, cytochrome c
            oxidase small subunit precursor (norC), NorF, cytochrome c
            oxidase polypeptide III-like protein (ntrE), calcium or iron-binding
            protein (tpc), NtrU (ntrU), and ferrienterobactin-like protein
            (fepA) genes, complete cds.
VERSION    AF083948
KEYWORDS   AF083948.1 GI:3925388
SOURCE     Pseudomonas sp. G-179
ORGANISM   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae.
REFERENCE  1 (bases 1 to 20761)
AUTHORS    Bedzyk,L., Wang,T. and Ye,R.W.
TITLE      The periplasmic nitrate reductase in Pseudomonas sp. strain G-179
            catalyzes the first step of denitrification
JOURNAL    J. Bacteriol. 181 (9), 2802-2806 (1999)
MEDLINE    99235760
PUBMED     10217771
2 (bases 1 to 20761)
AUTHORS    Ye,R.W., Bedzyk,L. and Wang,T.
TITLE      Direct Submission
JOURNAL    Submitted (12-AUG-1998) Life Sciences, DuPont, Experimental Station
            E328/148B, Wilmington, DE 19880, USA
FEATURES
  source
    1..20761
      /organism="Pseudomonas sp. G-179"
      /mol_type="genomic DNA"
      /strain="G-179"
      /db_xref="taxon:71238"
      /note="based on 16S rRNA gene sequence, this strain is
            similar to Rhizobium"
      /complement (519..1190)
      /gene="ntrR"
      /complement (519..1190)
      /gene="ntrR"
      /note="NtrR; putative regulator of nitrite and nitric
            oxide reduction"
      /codon_start=1
      /transl_table=11
      /product="FNR-type regulatory protein"
      /protein_id="AAC79437.1"
      /db_xref="GI:3925389"
      /translation="MPDELDRLQLQASARPAQDAVFEQGPADSFYLLHGRLLV
            TQVDEGQOIIIVVHGDLDGFAARPAQDAVFEQGPADSFYLLHGRLLV
            NPLATYAMQTIERLEAARVRIRREMSSTEVEERVAHTVIRLAKAKGKQKQKDIRIDF
            LISRODAEMTGTLTHTVSRILTSALMQLGVGQKQLFKVDISGLTTASGEFRKQ
            GTSQ"
  gene
  CDS

```

```

gene      complement (1331..2227)
CDS       /gene="ntrV"
          /complement (1331..2227)
          /gene="ntrV"
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /product="NtrV"
          /protein_id="AAC79438.1"
          /db_xref="GI:3925390"
          /translation="MTFRDAAFRRNVFLSTAPPALVNGLTGAIGFQSGLLPFRREG
            PVPAITISIPAGLLEIYESGFFRRNGLIVDAIPMTKIRHQSLLEIKFQVSGESYQCAAE
            GVCARDLSSGRADMPATGVSPDDAAAYARMLSNRTGEIWLRLPKLELAFAAGSKYP
            DDAIGIDPDSKNPALMLADYERETARKAFRISYPOPYGTGGENEGLADFGGNIWEM
            TSTCLRVTLDKRSKPISEAESCCGIYIATGKHGAPLSSFIREEKGGGCVGAPDNDVG
            FRLVRDPRWYTPILFALISGRSM"
          /complement (2309..3448)
          /gene="ntrU"
          /note="synonym: nirK"
          /complement (2309..3448)
          /gene="ntrU"
          /codon_start=1
          /transl_table=11
          /product="Cu-type dissimilatory nitrite reductase"
          /protein_id="AAC79439.1"
          /db_xref="GI:3925391"
          /translation="MSEORFLTRSRMLAGAAVAGALAPVVTVAHAEGGKIKTNSAAT
            AAIATLELVEKVEIYLPVPHATOKAEGEPKVEKMTIOEKIIVDDKGEVHAMT
            FDSVGPRMVIHQDDYVELTLVNPPTNLOHIDPHSATGALGGALTYVNGDTAV
            LRRKATKAGFVYHCAQAGVPEWHTVSGNKGALIMVLRDGLKDHKHELVLDVYVVG
            EODFYVQKENGKFKKYESAGEAYPDVLEAMKTLTPTHVFNAGAVAGLTDNALQAKV
            GDRVLLHSQANRDTDRPHLIGHGIDYVMAKFRANPELDQETWFIPGGAAGAAVYTF
            OORGIYAVVNHNLIEAFELGAAGHFVVTGDMNDLMTAVVSPTS"
          3911..4126
          /gene="napB"
          3911..4126
          /gene="napE"
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /product="NapB"
          /protein_id="AAC79440.1"
          /db_xref="GI:3925392"
          /translation="MAEPIATLEQTSFROKRRRELLTFVLAFGIWPVAVGVATYG
            FAVMAVOIYVGPGRPHDITPARNSAE"
          4132..4632
          /gene="napF"
          4132..4632
          /gene="napF"
          /note="NapF; putative electron transporter"
          /codon_start=1
          /transl_table=11
          /product="ferredoxin-like protein"
          /protein_id="AAC79441.1"
          /db_xref="GI:3925393"
          /translation="MGQVASTRAAFLTGLVBNGHYPPGVTDSEVSNCGCGDCVD
            ACPTKILSRDRIPITDFOVGCTFCGCSACQPEVPPEPATSPAHAVIADNCLIA
            LNPVQOACRDVCPPTTALRFKPRLGSPFPVLDTPDCTGCACISICPQOSITMTANT
            MEYAA"
          4625..4906
          /gene="napD"
          4625..4906
          /gene="napD"
          /codon_start=1
          /transl_table=11
          /product="NapD"
          /protein_id="AAC79442.1"
          /db_xref="GI:3925394"
          /translation="MREPOHHSNVVVVRLEAVERMTAALTEIEGEVHAAGCGKIV
            IYIEGSSSEFLPETILIRISAMDVIAAHNVFEGTIAEKMSDDGRTHAA"
          4881..7385
          /gene="napA"
          4881..7385
  CDS

```


QY 363 GlnThrAspAlaValTyrLeuPro-----GluGlyAlaProGln 375
 |||||
 Db 2335 -----GCCGTGGTTCCGCGACCTCGGGTTGACGGTGTCCGCCCGCGCGATCA 2285
 QY 376 AlaIleAspThrGlnGluAlaProLysThrProAlaProAlaLeuGlnGluGln 395
 2284 GGGCGCGCGCCACATCTTATGCGCGCGCGGAGCGGAGAGATCGAGATCCCATG 2225
 QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnGlyAla----- 407
 |||||
 Db 2224 ACCTTTCCGACGCGCTCTTCTTCCGATGTGTTTTCCTTCCTTCCGTTCCG 2166
 QY 408 AlaCysHisGlnProAspGlyLysGlyValProAsnAlaPheProLeuAlaAsnSer 427
 |||||
 Db 2165 GCTTGTGTGCGCTGACCGCGCG-----ATCGCGCTTCAGTCGCGACTTTTGCC 2115
 QY 428 AspTyrLeu-----AsnAlaAspHis-AlaArgAlaAlaSerIleVal 442
 |||||
 Db 2114 GACGCGCGAAGGCGCGCGCTCCGCGCGCGCATCAGATCCCGCGCGCTTCGCA 2055
 QY 442 AsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerValMetProAl 462
 |||||
 Db 2054 ATATCGAAGAAAGTGTGATGTTCTTCCGAAACGCTCTCACCGTCGACGCGCGATGACCA 1995
 QY 462 aile 463
 |||||
 Db 1994 AATT 1991
 RESULT 23
 AF051831 1709 bp DNA linear BCT 05-MAR-1999
 LOCUS Alcaligenes xylosoxidans dissimilatory copper-containing nitrite
 DEFINITION reductase (nir) gene, complete cds.
 ACCESSION AF051831
 VERSION AF051831.1 GI:2967826
 KEYWORDS
 SOURCE Achromobacter xylosoxidans
 ORGANISM Achromobacter xylosoxidans
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 Alcaligenaceae; Achromobacter.
 REFERENCE
 AUTHORS 1 (bases 1 to 1709)
 TITLE Suzuki, E., Horikoshi, N. and Kohzuma, T.
 Cloning, sequencing, and transcriptional studies of the gene
 encoding copper-containing nitrite reductase from *Alcaligenes*
xylosoxidans NCIMB 11015
 JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 427-431 (1999)
 MEDLINE 99160880
 PUBMED 10049725
 REFERENCE
 AUTHORS 2 (bases 1 to 1709)
 TITLE Suzuki, E., Horikoshi, N. and Kohzuma, T.
 Direct Submision
 JOURNAL Submitted (03-MAR-1998) Environmental Sciences, Ibaraki University,
 2-1-1 Bunkyo, Mito, Ibaraki 310-8512, Japan
 FEATURES
 source
 1. 1709
 /organism="Achromobacter xylosoxidans"
 /mol_type="genomic DNA"
 /strain="NCIMB11015"
 /db_xref="taxon:85698"
 gene
 348..>1662
 /gene="nir"
 misc_feature
 348..360
 /gene="nir"
 misc_feature
 435..462
 /note="fur-box"
 mRNA
 488..>1662
 /note="ntra-box"
 RBS
 568..573
 /gene="nir"

CDS
 580..1662
 /gene="nir"
 /codon_start=1
 /transl_table=11
 /product="dissimilatory copper-containing nitrite
 reductase"
 /db_xref="GI:2967827"
 /protein_id="AAC05831.1"
 /translation="MNALRPYLALAFMAAGTAAODADKLPHTKYTLVAPQVH
 PHEQATSSRPKVEFTMTTEKKMVIDDGGTTLQAMTFNSMGPTLVNHEGYVOLT
 LVNPAFNAMPHNVDFHGAATGALGATLVNNEBQVTLFPAKRSSTFYHCAPEGV
 PMHVSQMSGTLMLVPRDLKDPQKPLHYDRVTLGEFDLYIPKSGDYKQYATLA
 ESYGDVQVMRTLTPSHIVFNGVGLTANALTAIVAGTVLLIHSQARDPDPHILG
 GHGDVMEWGKEFANPQPDLETFIRGSGAAGALYFKQPGVAYALNHLIEAFELGA
 AGHIVKEGKMDIMKQIKAPAPIPR"
 BASE COUNT 333 a 577 c 560 g 239 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,37e-21 Length: 1709
 Score: 452.00 Matches: 135
 Percent Similarity: 49.73% Conservative: 49
 Best Local Similarity: 36.49% Mismatches: 142
 Query Match: 17.21% Indels: 44
 DB: 1 Gaps: 16
 US-10-088-045-2 (1-502) x AF051831 (1-1709)
 QY 41 AlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGluGlnLeuPro 60
 |||||
 Db 631 GCGCGCGGAGCCCGCTGCGCGCGCGAGCGCC-----GACACGCTGCC 672
 QY 61 ValIleAspAlaIleValThrHisAlaProGluValProPro-----ProValaParg 78
 |||||
 Db 673 CATACCAAGGTACGCTGTGCTCCGCCCGCGCGCATCTCCGACAGACGCGACCAAG 732
 QY 79 AspHisProAlaLysValValValLysMetGluThrValGluLysValMetArgLeuAla 98
 |||||
 Db 733 TCGCGCGCCAAAGTGTGATGTTCTTCCGAAACGCTCTCACCGTCGACGCGCGATG 792
 QY 99 Asp--GlyValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIle 117
 |||||
 Db 793 GACAAAGGACGACGCTGACGCGCATGACGTTCAAGGCTCCATGCCCGCGCGCGCTG 852
 QY 118 ArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMet 137
 |||||
 Db 853 GTGCTCAGAGGCGCATGCTGACGCTGACGCTGCTCAACCGCGCACCAAGCCCATG 912
 QY 138 ProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluAlaSerPhe 157
 |||||
 Db 913 CGCACAAAGTGTGATGTTCTTCCGAAACGCTCTCACCGTCGACGCGCGATG 972
 QY 158 ThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrVal 177
 |||||
 Db 973 GTCAACCTGCGACGACGCTGACGCTGCTCAAGCGCGACGCGCATCTTCGTC 1032
 QY 178 TyrHisCysAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMetTyr 194
 |||||
 Db 1033 TACCACTGC-----CGCCCGCAAGGATGTCGCTGCGACGCTGTGCGGACATGAC 1086
 QY 195 GlyLeuIleLeuValGluProLysGluGlyLeuPro-----Lys 207
 |||||
 Db 1087 GGCAGCGTATGCTGTGCGCGCGCGCGCTGAAAGATCCGACGCGCAAGCGCGCAT 1146
 QY 208 ValAspLysGluTyrTyrValMetGlnGlyAspPheTyr-----ThrLysGly 223
 |||||
 Db 1147 TACGACCGCGCTACACATCGCGAGTTCGACTGTATATCCCAAGGCGCGCGACG 1206
 QY 224 LysTyrGlyGluGlnGlyLeu-----GlnProPheAspMetLysAlaIleArg 240
 |||||
 Db 1207 AAGTACAGACATCAACCGCGCTGCGCAAGCTATGCGACGCGTACGATGCGCG 1266
 QY 241 Glu--AspAlaGluTyrValValPheAsnGlySerValGlyValaLeuThrGlyGluAsn 259


```

Db      598 ATCGGCGAGTCGACCTGTATATCCCAAGACAGCGCCACTACAGACTACCG 657
Qy      233 Pheaphecglyalailearglu-----AspAlaGlu 244
Db      658 ---GACCTGGGCTCAGTTACAGACACCCCGGTCATGCCACCTGACCCGAGC 714
Qy      245 TyrValValPheAenglySerValGlyAlaLeuThrglyUaenAlaLeuLysAlaLys 264
Db      715 CACGTGGTGTTCACAGCGCGGTGTGCGCGCTGACCGGCGCCAGCGGCTGACCTCCAG 774
Qy      265 ValGlyLthrValAlaGleuPheValGlyAenGlyProAenLeuThrSerPhe 284
Db      775 GTCCGGAGAGCGG---CTGTTCAATC---CACGCCAGGCCAACCGCAGCGCTGCC 828
Qy      285 HisValIleeglygluilePheApLyValHisPhegluGlyLysGlyLys--- 303
Db      829 CACCTGATTGGCGGCCACGGGACTGGGTC---TGGACACCGGCCAGTTGCCAACCG 885
Qy      304 -----HisaniIeginThrThrLeuIleProAlaGlyAlaAlaIleThrgluPhe 321
Db      886 CCGCAGCAACATGAAACCTGGTTATCCCTCGCGGTCTGCGGTGGCGCGCTGTAC 945
Qy      322 LysValAlaPheValProGlyAsePtyrValLeuValAlaPheSalAlaIlePheAlaPheAsn 341
Db      946 ACCTTCAGACGACCGGGGACCTACGTGATCCTGACCATTAACATGAGAGCCATGAA 1005
Qy      342 LysGlyAlaLeuGlyLthrValGlyLysValGlyLysGluGluAenHisGluIle 358
Db      1006 CTGGGGGCTCTGGCCCAATCAATGATGAGGGGAGGAGGAGCATGACGACTG 1056

RESULT 25
AP005961 298550 bp DNA linear BCT 28-MAR-2003
LOCUS     Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
DEFINITION
ACCESSION AP005961 BA000040
VERSION    AP005961.1 GI:27355365
KEYWORDS
SOURCE     Bradyrhizobium japonicum USDA 110
ORGANISM   Bradyrhizobium japonicum
REFERENCE  1
AUTHORS    Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiyumi,T.,
            Sasemoto,S., Watanabe,A., Ideesawa,K., Iriuchih,M., Kawashima,K.,
            Kohara,M., Matsumoto,M., Shimo,S., Tsuruoka,H., Wada,T., Yamada,M.
            and Tabata,S.
            Complete genomic sequence of nitrogen-fixing symbiotic bacterium
            Bradyrhizobium japonicum USDA110
            DNA Res. 9 (6), 189-197 (2002)
JOURNAL    22484998
MEDLINE    12597275
PUBMED     12597275
REFERENCE  2
AUTHORS    Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiyumi,T.,
            Sasemoto,S., Watanabe,A., Ideesawa,K., Iriuchih,M., Kawashima,K.,
            Kohara,M., Matsumoto,M., Shimo,S., Tsuruoka,H., Wada,T., Yamada,M.
            and Tabata,S.
            Complete genomic sequence of nitrogen-fixing symbiotic bacterium
            Bradyrhizobium japonicum USDA110 (supplement)
            DNA Res. 9 (6), 225-256 (2002)
JOURNAL    22485002
MEDLINE    12597279
PUBMED     12597279
REFERENCE  3
AUTHORS    Kaneko,T.
            Direct Submission
            Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research, 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail: kaneko@kazusa.or.jp,
            URL: http://www.kazusa.or.jp/rhizobase/
            Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
            Location/Qualifiers
            1. 298550
FEATURES
Source

```

```

/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
11..721
/gene="nmrR"
11..721
/gene="nmrR"
/feature="Cp family
ORF ID: b1r7084"
/codon_start=1
/transl_table=11
/product="transcriptional regulatory protein"
/protein_id="BAC52349.1"
/db_xref="GI:27355366"
/translation="MAKUTSUYAHU.PI.PAGYTPALDELTPARSAPKNSAIFEO
GADQSFLLHCHVRAAKTTPTGEQIVRYVAPGTFGLAAMIGRYPATMAVD
SVVLTPSAMPRLVPRPALANTQTGTQLQESRLTQEWSTQVEQRVIAHALR
LAQSKKJLDHGIIDPISRODIAQMTGTLTHVSRILSGWEOGLVBSGRORILR
EPHRIIVLERSPSGAP"
complement (842..1039)
/gene="b1r7085"
complement (842..1039)

```

```

gene
CDS
/gene="b1r7085"
complement (842..1039)
/gene="b1r7085"
/feature="ORF ID: b1r7085
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC52350.1"
/db_xref="GI:27355367"
/translation="MPFGSDLVDDIMRTAHTIRVFLAFLACVGCPIATFTVEDA
CREHGIDRDKFLALGDCVPA"
complement (1297..2649)
/gene="hemN"
complement (1297..2649)

```

```

gene
CDS
/gene="hemN"
/feature="ORF ID: b1r7086"
/codon_start=1
/transl_table=11
/product="anaerobic coproporphyrinogen III oxidase"
/protein_id="BAC52351.1"
/db_xref="GI:27355368"
/translation="MRADLAVSYGEERLPRYTSYPAPHPSPVIDAGYARWLSEPA
GASASLYLWPCFRCMCWYCGCTQVRRDDILIAAYORTLRSEIALVAETIGRIKVE
HIHFGGPTPIAPEAFELMAAMRQAFVLPSEAIJAVEIDPTTLADPAVEMRLSGV
NMSAGVQSFDPVQVATINRISFEGQTAIVVMILBRAGIAGINFDLIYGI.PHOTVASC
LDTVRSLLAPDRSVFGYAHVPDKKIQRMNQCALPDGPARHQAICAIYALXEA
GVYQIGLDHFAARDSDMAVAFERTLRRNFQGTIDQGEVLGFGSALGILPQGYVQ
NEVOIGLVAQSIGASRLATAGYGLTDDRLRADIIERINCEFSADLGICAHGAEF
EAMLKASRLKPLISGVVRLDGDRLAVNDSFVRSVAAPDAHLDPGKQLHSBAY"

```

```

gene
CDS
/gene="b1r7087"
/feature="ORF ID: b1r7087
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC52352.1"
/db_xref="GI:27355369"
/translation="MRLALELILFSLLDVGRGHI.PRPDPFPVPASAPAGIVHTLAAV
IVYLAIVLSAIVAAVVAIHL"
3079..3954
/gene="b1r7088"
3079..3954
/gene="b1r7088"
/feature="ORF ID: b1r7088
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC52353.1"
/db_xref="GI:27355370"

```



```

Db      4747 GACACCCAGAGATGATGAAAGAACTGATCCATCCATCGTGTTCACGCGAAGTCC 4806
      253 G1VAlaLeuThrglyGluAaAlaLeuValaValaGlyGluThraValaArgLeuPhe 272
      4807 GGGCGCTTACCGGCAAGATGCTCCCTCCCAATGTGGCGAAGCGT-----CTG 4860
      273 ValGlyAaGlyGlyProAaLeuThrsSerPheHisValIleGlyGluIlePheAsp 292
      4861 ATCGTGACCTGCGAGCAATCGGACGCCGCCGATGTGATCGGCGCATGGGCGAC 4920
      293 LysValHisPheGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 309
      4921 TATGTC---TGGAGACCGCGCAATTCGCGCAATCACCGAGTCCGCGTGAAGCTGG 4977
      310 LeuIleProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 329
      4978 TTCAATCCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 5037
      330 ValLeuValaPheHisAlaIlePheArgAlaPheAaLeuGlyAlaLeuGlyIleLeuLys 349
      5038 GCGTATGTCAAGCATTAACCTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACAG 5097
      350 ValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 369
      5098 GTCGAAGCAAGTGAAGCAAGCATGTGATGACCGAG-----GTGAAAGCGC 5142
      370 ProGlyGlyAlaProGlyAlaIleAsp-----ThrGlnGluAlaProLysThrProAla 387
      5143 CTTGCGGAATCCCGCGCAACCACTAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5202
      388 ProAlaAaLeu-----GlnGlyGlnIleLysAlaGlyLysAlaThrTyAspSerAa 405
      5203 CTTTCTTC---CTTTCGGGGAGACGATCTGATGATCAAGTGAAGTGGTCTGGCT 5261
      406 CysAlaIaLysHisGlnProAaPheGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 425
      5262 TGGCGCGCC-----GGACTCGAGAGCGGATCGCGCGCGCGCGCGCGCGCTGTC 5306
      426 AaSerAaPheLys-----AaAlaAaPheHisAlaIleAlaIleAlaAa 443
      5307 TCGGACATGGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5366
      444 GlyLeuSer-----GlyLysIleThrValaGlnGlyAaGlnTyGlySerVal 459
      5367 AATCTGCTTATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5426
      460 MetProAla-----IleAlaLeuSer 466
      5427 CTGCGCGCGGATGCGGTTACAGAGCGCGCTGACGTGATCGCGCATCAGTTTCGTATCC 5486
      467 AspGlnGlnIle 470
      5487 GACTACCACTT 5498

```

RESULT 26
 LOCUS RHU65658
 DEFINITION Rhizobium hedysari copper containing nitrite reductase (nirK) gene,
 complete cds.
 ACCESSION U65658.1 GI:1488171
 VERSION 065658.1
 KEYWORDS Rhizobium sputiae
 SOURCE Rhizobium sputiae
 ORGANISM Rhizobium sputiae
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 Toffanin,A., Wu,Q., Mankus,M., Casella,S., Abruna,H.D. and
 Shapleigh,J.P.
 TITLE Characterization of the gene encoding nitrite reductase and the
 physiological consequences of its expression in the
 non-dentifying Rhizobium hedysari HCNMT1

JOURNAL Appl. Environ. Microbiol. (1996) In press
 REFERENCE 2 (bases 1 to 1499)
 AUTHORS Toffanin,A. and Shapleigh,J.P.
 TITLE Direct Submission
 JOURNAL Submitted (31-JUL-1996) Microbiology, Cornell University, Wing
 Hall, Ithaca, NY 14853, USA
 FEATURES
 source
 1. 1499
 /organism="Rhizobium sputiae"
 /mol_type="genomic DNA"
 /strain="HCNMT1"
 /db_xref="taxon:50338"
 40..52
 /note="similar to Fur binding motif in Escherichia coli"
 protein_bind
 40..52
 /note="similar to Fur binding motif in Escherichia coli"
 gene
 119..1264
 /gene="nirK"
 RBS
 119..122
 /gene="nirK"
 CDS
 131..1264
 /gene="nirK"
 /codon_start=1
 /transl_table=11
 /product="copper containing nitrite reductase"
 /protein_id="AB05880.1"
 /db_xref="GI:1488172"
 /translation="MTNTLQMPRTMLTGAAGALPTILTSGGNASPTVRKLSAT
 EIALPRRLDLYKPPFVYVHOKAGSGKPVVEVLTIEKKIVIDCKEYVNAFDP
 GSVKPLIVVHODIVETVLVNDPTTLTGNIDFHSATGLGKGLTVNPGSVAIR
 FKATKGVFVYHAPRGWVPMVHTSGMNAIYVLPREGALDGGKLVYDQVYLEQ
 DFYIPRDEKGEFKYDPSGEAYEDVAVNRKTLTPHIVNGAVGLTGSALTAAYGE
 RVLIVSQRNDRPRLHIGSGEYVVRKTFVVPDQDFTFIPPTGAAVAYTFEQ
 PGIVAVNNHLEAFELGAHAKVYGMNDMLMTVRSBSGS"
 BASE COUNT 323 a 443 c 428 g 305 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1499
 Score: 440.00
 Percent Similarity: 44.34%
 Best Local Similarity: 30.54%
 Query Match: 16.76%
 DB: 1
 Gaps: 15
 US-10-088-045-2 (1-502) x RHU65658 (1-1499)
 QY 1 MetSerLysProThrIleuIleuLysThrIleuIleCys---AlaLeuSerAlaLeuMet 19
 Db 149 ATGACCCGGCGCGAGTGTGACGCGCGCTGTGTCGCGAGCGCTGACCCCATCTTC 208
 QY 20 LeuSerGlyCysSerAaGlnAlaAaPheLysAlaAlaGlnProLysSerThrValaAsp 39
 Db 209 ACCTCGGGGGGGGGAATGCTCCCTCCCTGTCGAGAAAGCTCAGCGCATGAAATC 268
 QY 40 AlaAlaIaLysThrIlaAaAlaAaPheAaAlaIaSerGlnGlnIleGlnGlyLeu 59
 Db 269 GCTGCGCTTCCCGCGAGAACTGAGC----- 295
 QY 60 ProValIleAaPheAlaIleValThrHisAlaProGluValProProValaAaPheAsp 79
 Db 296 -----CTGTAAGCACTTTCGTCAAGTA 322
 QY 80 HisPro-----AlaLysValaValaValaLysMetGluThrValaGluLys 93
 Db 323 CATACCGAAAAGCAGAAAGCGGACCGAAGTGTGAAAGTCACTACATGATGAGAA 362
 QY 94 ValMetArgLeuAlaAsp-----GlyValGluTyGlnPheThrPheGlyGlyGln 111
 Db 383 AAGAACTGTGATGATGAGCAAGCGCAAGTAAGTCAAGCCATTTGACGATCG 442
 QY 112 ValProGlyGlnMetIleArgValaArgGlyGlyAspThrIleGluValaGlnPheSerAa 131
 Db 443 GTTCTGTGTCGCTCATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 502

QY 93 ---LysValMetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlyGln 111
 DB 1093 AAGAAAGGTCGATCGACGAGAGGATACCACTTCAGGGGATGACCTTCAAGGGCTCG 1152
 QY 112 ValProGlyMetLeuValArgGluGlyAspThrIleGluValGlnPheSerAsn 131
 DB 1153 ATGCCGGGCGCGTGTGCTGATGAGGACATGACCTAGCAAGCAACGACCGCTGTCAT 1212
 QY 132 HisProAspSerLysMetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGly 151
 DB 1213 CCGCGACCAACACATGCGGACACATGACCTTCACTCCGCGACGGGCGCTCGGC 1272
 QY 152 GlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheValLeu 171
 DB 1273 GGGGGCGCGCTCACCTGATCATCCGCGACAGGCGTGGTGGCTGGAGAGCGAC 1332
 QY 172 GlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGly-----Met 187
 DB 1333 AAGACCGGCGGTCTGCTATCACTGC-----GCGCGGGCGCGCGCATGATCCCTCG 1386
 QY 188 HisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGluProGlyGlyLeuPro--- 206
 DB 1387 CAGCTGCTCTCGGATGACGAGGTCGCGTGGTGGTGGTGGCGCGACGACTAAGCAG 1446
 QY 207 -----LysValAspLysGlyTyrTyrValMetGlnGlyAspPheTyr 220
 DB 1447 GGCAAGGCGCACGGCGTGAATATGACAGGCTACTATGTCGGCGACGAGCATATAC 1506
 QY 221 Thr-----LysGly-----LysTyrGlyGlnGlyLeuGlnProPhe 233
 DB 1507 GTGCCCGCGACGAGAGGCAATTTCAAGTCTATGACTCGCGCGCGAAGCTTCCACC 1566
 QY 234 AspMetGlyLysValAlaIleArgGlu---AspAlaGluTyrValAlaPheAsnGlySerVal 252
 DB 1567 GACACCGAAGAGATGAGAGAGACCTGATCCCATGCTGGTTCACAGGAGAGCTC 1626
 QY 253 GlyAlaLeuThrGlyGluAsnAlaLeuLysAlaLysValGlyIleThrValArgLeuPhe 272
 DB 1627 GGGCGGCTTACCGGCAAGATGGCTCACCGCAATGTCGGCGAGAACTG-----CTG 1680
 QY 273 ValGlyAsnGlyTyrProAsnLeuThrSerSerPheHisValIleGlyIlePheAsp 292
 DB 1681 ATCGTGACTTCGACGGCCCAATCGGACAGCGCGCGCATGATCGCGCGCATGAGCGAC 1740
 QY 293 LysValHisPheGlyGlyGlyGlyGluAsnHis-----AsnIleGlnThrThr 309
 DB 1741 TATGTC--TGGGAGACCGGCAAAATTGGGCAATCCACCGAGGTCGGGCTGGAGACTCG 1797
 QY 310 LeuIleProAlaGlyAlaAlaAlaIleThrGluPheLysValAlaValProGlyAspTyr 329
 DB 1798 TTATCCGGCGGCGCTGGGACGGGCGTCGATGACAAATTCAATGACGACCGCGATATAC 1857
 QY 330 ValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaGlyIleLeuLys 349
 DB 1858 GCGTATGTCACGCTAATCGATGAGGCGCGCATCTGCGCACCGCGCATCTTCAAG 1917
 QY 350 ValGluGlyGluGluAsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyrLeu 369
 DB 1918 GTGGAGGCAAGTGGAGAGAGATGATGACCGAGGTGAAGCGGCTCGG----- 1968
 QY 370 ProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAla 389
 DB 1969 -----GAAATCCCGCGCAACACC 1986
 QY 390 AsnLeuGlnGluGlnIleLysAlaGlyLysAlaThrTyrAspSerAsnCysAlaAlaCys 409
 DB 1987 AACTAGACGCGACAA--GGGCGCGG----- 2010
 QY 410 HisGlnProAspGlyLysGlyValProAsnAlaPheProProLeuAlaAsnSerAspTyr 429
 DB 2011 -----GGAATCACCGGCGGCTTCTTCTTGGGGGAGAGCATG 2049

QY 430 LeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIle 449
 DB 2050 CTGATGCACTTCAAGCTGAAGATGTCCTGCTGGCGCGCGCATGCGACGACGATC 2109
 QY 450 ThrVal 451
 DB 2110 GCGGTG 2115
 RESULT 28
 AE009732/c 11630 bp DNA linear BCT 20-MAR-2003
 LOCUS Brucella melitensis 16M chromosome II, section 94 of 107 of the complete sequence.
 DEFINITION AE009732 AE008918
 ACCESSION AE009732.1 GI:17985201
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Brucella melitensis 16M
 Brucella melitensis 16M
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Brucellaceae; Brucella.
 1 (bases 1 to 11630)
 DelVecchio, V.G., Kapral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goldsman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.-J., Haselkorn, R., Kyriides, N. and Overbeek, R.
 The genome sequence of the facultative intracellular pathogen *Brucella melitensis* Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
 2 (bases 1 to 11630)
 DelVecchio, V.G., Redkar, R.J., Patra, G. and Mujer, C.
 Direct Submission
 Submitted (13-NOV-2001) Institute of Molecular Biology and Medicine, University of Scranton, Scranton, PA 18510, USA
 3 (bases 1 to 11630)
 Elzer, P.H. and Hagius, S.
 Direct Submission
 Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA
 4 (bases 1 to 11630)
 Kapral, V., Los, T., Ivanova, N., Anderson, I., Bhattacharya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goldsman, E., Selkov, E., Haselkorn, R., Kyriides, N. and Overbeek, R.
 Direct Submission
 Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell Park Drive, IL 60612, USA
 5 (bases 1 to 11630)
 Letesson, J.-J.
 Direct Submission
 Submitted (13-NOV-2001) Unite de Recherche en Biologie Molculaire, Laboratoire d'Immunologie et de Microbiologie, Universite of Namur, 61 rue de Bruxelles, Namur 5000, Belgium
 6 (bases 1 to 11630)
 O'Callaghan, D.
 Direct Submission
 Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue Kennedy, Nimes 30900, France
 Location/Qualifiers
 1. 11630
 /organism="Brucella melitensis 16M"
 /mol_type="genomic DNA"
 /strain="16M"
 /db_xref="taxon:224914"
 /chromosome="II"
 /complement(132..1262)
 /gene="BMEII0988"
 /complement(132..1262)
 /gene="BMEII0988"
 /BC_number="1.7.99.3"
 /codon_start=1

gene
CDS
/transl_table=11
/product="COPPER-CONTAINING NITRITE REDUCTASE PRECURSOR"
/protein_id="AAL54230.1"
/db_xref="GI:17985202"
/translation="MADQIOVNRRTILAGALAGPULSATSGCGRKAAAE
IALPRKVELVPPFVPHATQVABGSPKVTQWTVEKKYIIDAGTEVHAMTRG
TVGSLPMVHQDDYLETILNPTNTLLHIDPHATGALGGGLTEINKEGTVALE
KATKPGVFNHCAFGVPMWVSGNAGVAVLPRGLHDGKNKLTVDKVVYVGGD
FYVPRDNGNKTVEAAGDAEDTVKMTLTPTHVFNAGVALLTGDKALTPAKGSK
VLIIHSDANDTRPHLIGHDYVMATGKFTPPDVOETWTFPGGAAAAAFTFRQP
GIYAVYNHNLIEAFELGAAPFKVTGEMNDLMTSLAPSGM"
/complement (1618. .2220)
/gene="BMEI10989"
/complement (1618. .2220)
/gene="BMEI10989"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL54231.1"
/db_xref="GI:17985203"
/translation="MGLLAGALVLPSPAPAKTQQAAMPNATSPHQADVLLRGA
DIRSTGIDEGAEIAGVAVNAGVAMRLVNRIVADQCKGHLPPVLIIGSLGN
AATVIAEELERBGIAVDYMAFTAGTGPDLPGVVRVYVNFYFKHGKGLPLVGGPRH
GHLENDPSNAKVGAFENIEKQRPLOAEVVDVLAVERNAD"
/complement (2267. .2851)
/gene="BMEI10990"
/complement (2267. .2851)
/gene="BMEI10990"
/codon_start=1
/transl_table=11
/product="farnesyl cysteine carboxyl-methyltransferase"
/protein_id="AAL54232.1"
/db_xref="GI:17985204"
/translation="MMPLIALIWSIGLVAVVIRIPHORARKIKVAPARTLADRL
ALCAATVGLSVPLVYATGFPKADTFPPMGMIGTLELFLMFLFYSHQKGN
WSYSLDIRREHKLVTDLRYRHPMLTSLWLAIDFLLPMWLAGLGLLVALIY
FYRIEHEAMRAFGSAYDESSRIGRIIPKWM"
/complement (2848. .4023)
/gene="BMEI10991"
/complement (2848. .4023)
/gene="BMEI10991"
/codon_start=1
/transl_table=11
/product="Hypothetical Membrane Spanning Protein"
/protein_id="AAL54233.1"
/db_xref="GI:17985205"
/translation="MQHPVRRRAIFPGYDPTPEAFENRLLREIARFALMDVKTO
VSPVSEASSSEIGSVITETRASGWSQTDVFNFLVDKIVLAFARPLPRLMKYLA
FADVLSGTAFFRFARARFALYFLYFPLVLFALAGMLAARLTAPWLAGMIVGL
FVFPLIATLTGNMSTNHLMDLMSFSNFRGQRPDAELMORFAEIIDRVAAGHD
EVIIVIGSTGMLMDVVARCLADRAFATRAPKVVLLTGSTALAKAGYHPAAGKPD
DNORLVDPEALCEVOCLTDVNFYTPDVAEMKLRPADRAFPVIVROIRIDMOP
DIYRIKRNFRHYQVFGNTKAYMTDFQICGPAISLSERMMDLVGTAKESAQQ"
/complement (4147. .4473)
/gene="BMEI10992"
/complement (4147. .4473)
/gene="BMEI10992"
/codon_start=1
/transl_table=11
/product="WRP PROTEIN HOMOLOG A"
/protein_id="AAL54234.1"
/db_xref="GI:17985206"
/translation="MEBKOLERIKACVREALRAVODPELAVNLVDGMIVAVEISDQ
MNVKMTTTRGCPAGFLQAVQACIEIEIGTGAARVELTYRPEMKPEMAIEVQAI
FAAPSP"
/complement (4480. .5007)
/gene="BMEI10993"
/complement (4480. .5007)
/gene="BMEI10993"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL54235.1"

gene
CDS
/db_xref="GI:17985207"
/translation="MTVOELAPYELDLRPLKSGVPEFQALNGAVDSLAPGGQLLIA
PFKPVLLFVLEKRGFEAKPEPIGEGDMVLSPVPEPLAEVQSLSNVGSPODIWPDPL
OYLDCTDMDPEPELVIRILAEVENMSAGEVLFALHREPLFLPELENRQHGWGNFDE
TKQTVRIMRVGGCG"
/complement (5004. .5840)
/gene="BMEI10994"
/complement (5004. .5840)
/gene="BMEI10994"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL54236.1"
/db_xref="GI:17985208"
/translation="MLNITAGLSAPPIHAALGIGMMFTFAMGVSYRLTWFLSPDE
RRTTRLVMMAGAAALTVLAGVAFAGQDMADVLLAALLPGVASVLYAFPMRAVY
ROKRRKATLINSASIPAPAAVIALILALAPMCPSPMIGALVYLPFGMLTGS
LMOYKIVFPLTLECYGVMGKVPFTRQDIYSEKPARMFCITYTGVALATLGA
GYPVAFQLASALNLIILALTVHFFRARRLMDVPEAVRLNGVALPHLYAGGPVPR
SK"
/complement (6407. .6721)
/gene="BMEI10995"
/complement (6407. .6721)
/gene="BMEI10995"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL54237.1"
/db_xref="GI:17985209"
/translation="MDGPRARQVNLFPFPLSSTPAQFPPTTHATIFSITDLE
PGSEPIVNDHDPVPLRNQLEHRNPGAFSVMYKETGPLRVEIGRRKVASDHECTG
NH"
/complement (6878. .8779)
/gene="BMEI10996"
/complement (6878. .8779)
/gene="BMEI10996"
/codon_start=1
/transl_table=11
/product="NORD protein"
/protein_id="AAL54238.1"
/db_xref="GI:17985210"
/translation="MLDPLELEETVGRANRHLIGTGSNPOYDHAVALVDIQRILAI
CFRFGGDIIVQVAPARARTSHRLGRMALGEBKLOPLDEATVLMIPPIALFP
DROLNDLYLVIVGVAWMPDADLPELRLRDLAALQIAQTVRACAPFGLKPR
YKRLCAAILABRPKRLHRLQOQVEARLISLKGQADLPDAPLPTIFPHGAPGYLA
LPVLPMPGKMKREVA PRGCEDEPVNQSSEGAETROIAQREORPRHADSPFTLN
RKETILAMAMSVDRPSGSDONAKSADLDDLTUGERKGRPARFRFDLDPBEA
VRSLLTAELTYPEMDYRGAVALPNHCAVLAAGVQKRPLEIDAAQSLVRVRQF
EILRRPEREVLRLQOLDGTDLDLDAVVRSCDLAAGGQSDRVHLSRPOANDLVTLLV
DVSIGTDVANDRRKVDVEKEALLVLANGIAACGDRCSILTFSTRRSWVRVTVDF
DESFGTVEHRIIAALRGFYTTRGAMRHATKLABOPRKKLLLLTGCKPMDVDHY
EGRRALDEBRRAAGFRAKGVNVAFTVDRBSAYIPALFGRGYALVANLALPVAL
PATRYVLAC"
/complement (8791. .9603)
/gene="BMEI10997"
/complement (8791. .9603)
/gene="BMEI10997"
/codon_start=1
/transl_table=11
/product="NOKO PROTEIN"
/protein_id="AAL54239.1"
/db_xref="GI:17985211"
/translation="MMSLKAVETASPAIPAYSPGNECALFEMAMORKLPLLLKQPT
GCKKTRFVSHMARMOPLSTVTSCHDDLAADLTGRYLLKGGDTIMMDGLTFAVADG

Alignment Scores:
Pred. No.: 5.33e-19 Length: 11630
Score: 431.00 Matches: 122
Percent Similarity: 48.43% Conservative: 48
Best Local Similarity: 34.76% Mismatches: 147
Query Match: 16.41% Indels: 34
DB: 1 Gaps: 11

US-10-088-045-2 (1-502) x AE009732 (1-11630)

43 LysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnIleGlnGluProValIle 62
 1151 AGAAGGCGGACCGGAGGAAATTTCTGCTCCGAGACAG---AAAGTGAGCTGTC 1095
 63 AspAlaIleValThrHisAlaProGluValProProValAlaAspArgHisProAla 82
 1094 GATCTCCCTTCCTGTCATGCA-----CACACTCAGGTTGGCGAAGCGCGCGAAG 1044
 83 LysValValValLysMetGluThrValGlnLysValMetArgLysAlaAsp--GlyVal 101
 1043 GTGGTTAGTTCACCATGCGTATGAGAGAAAGAAAGATCGTATGATGATGCGGACG 984
 102 GluTyrGlnPheThrPheGlnGlyGlnValProGlnGlnMetIleArgValArgGlu 121
 983 GAACTCATGCGCATGATTTCAACGGAGCGGTGCGGACCGCTCATGCTGTCATCAG 924
 122 GlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsnVal 141
 923 GAGCATATCTCGAACTGACGCTCATTAACCGGAAACCATATACCTGTTGCAACATC 864
 142 AspPheHisAlaAlaThrGlyProGlyGlyValaGluAlaSerPheThrAlaProGly 161
 863 GATTTTCATCGCGGACCGGACGCTGCGGCGGCTGCGGAGATCAATCCGCGC 804
 162 HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHisCysAla 181
 803 GAGAAAGCCGTCCTCGGCTTCNAAGCCAAACCTGCGCTTGTGCTCATCTGTC--- 747
 182 ValAlaProValGlyMet-----HisIleAlaAsnGlyMetTyrGlyLeuIleu 198
 746 ---GCACCTCCCGGAATGTTCCGTCGACGTCGTATCGGCGCATGAACGGCGCTCATG 690
 199 ValGluProLysGlnGlyLeuPro-----LysValAspLysGlu 211
 689 GTTCTGCGCGCGGCGGCTGATGACGAGCGAAGGCAACGCTGACCTTATGACAAAGTC 630
 212 TyrTyrValMetGlnGlyAspPheTyrThrLys-----GlyLys 224
 629 TATTAATGTCGCGGACGAGATTTTATATGTCGCGGTATGAGAAAGCGCAATTACAGACC 570
 225 TyrGlyGlnGlnGlyLeuGlnProPheAspMetGlnLysAlaIleArgGlu--AspAla 243
 569 TATGAAGCGCGCGGCGATGCTATGAGATACGCTCAAGGATGCGCACCTGACGCGC 510
 244 GluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAla 263
 509 ACCCATGTGTCCTCAATGTCGTCGTCCTTACGCGGTGACAAAGCCCTGACGCGC 450
 264 LysValGlyGluThrValArgLysPheValGlyAsnGlyGlyProAsnLeuThrSerSer 283
 449 AAAGTCGCGGAAAGGTCGATGATC-----CACTCGCAAGCGCAACCGCGATACCGGT 336
 284 PheHisValIleGlyGlnIlePheAspLysValHisPheGlnGlyLysGlyLysAsn 303
 395 CCGCACTGATCGCGCGGCGATGCGATTAATGTCGCGGACCGCAAGTCAATACGCGC 336
 304 HisAsnIle-----GlnThrThrLeuIleProAlaGlyAlaAlaIleThrGluPhe 321
 335 CCAGATGCTCAGCAGAAACCTGTTATTCCTCGGCGCGCGGCGGAGCGCTTCTAC 276
 322 LysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsn 341
 275 ACCTTCGCGGACCGGCGATCTACGCTATGTCATATTAACCTGATGAAAGCTTTGAG 216
 342 LysGlyAlaLeuGlyLysLeuLysValGlyGlyGlnLysAsnHisGlnIleTyrSerHis 361
 215 CTGGGCGCGGACCGCATCTCAAGTGAAGCGGGAATGAGATGACACTTGATGAC--- 159
 362 LysGlnThrAspAlaValTyrLeuProGlnGly 372
 158 -----TCGATACTGCGCGCTCGCGG 138

RESULT 29

AE014526
 LOCUS

AE014526 11302 bp DNA linear BCT 02-OCT-2002
 Brucella suis 1330 chromosome II section 22 of 108 of the complete
 genome.

DEFINITION

AE014526 AE014292
 AE014526.1 GI:23463593

ACCESSION

Brucella suis 1330

VERSION

Brucella suis 1330

KEYWORDS

Brucella suis 1330

SOURCE

Brucella suis 1330

ORGANISM

Brucella suis 1330

REFERENCE

Brucella suis 1330

AUTHORS

Brucella suis 1330

REFERENCE

Brucella suis 1330

AUTHORS

Brucella suis 1330

TITLE

Brucella suis 1330

JOURNAL

Brucella suis 1330

FEATURES

Brucella suis 1330

source

Brucella suis 1330

gene

Brucella suis 1330

CDS

Brucella suis 1330

gene

Brucella suis 1330

CDS

Brucella suis 1330

gene

Brucella suis 1330

CDS

Brucella suis 1330

gene

Brucella suis 1330

CDS

Brucella suis 1330

gene

Brucella suis 1330

CDS

Brucella suis 1330

gene

Brucella suis 1330

CDS

Brucella suis 1330


```

Db      5718 GATCTCCCTTCGTCATGCA-----CACACTCAGGTGGCGAAGCGCGCCGAG 5768
      83 LysValValValLysMetGluThrValGluLysValMetArgLysLeuAAsp---GlyVal 101
      5769 GTGGTTCAGTTCACCATGTCATGTCAGAGAAAGAAAGATTCATGATGCGGGGACG 5828
      102 GluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIleAspValArgGlu 121
      5829 GAAGTCATCCATGATGATTCATTCACGAGACGTCGCCGACCGCTCGTGTGTCATCAG 5888
      122 GlyAspThrIleGluValGlnPheSerAsnIleProAspSerLysMetProIleAsnVal 141
      5889 GACGATATCTCGAATCGACGCTCATTAACCCGAAACCAATACCTGTTGCACAAATC 5948
      142 AspPheIleAlaIleAlaThrGlyProGlyGlyGlyValAspLysAspPheThrAlaProGly 161
      5949 GATTTTCATCGCGCCACCGCGCATCGGGTGGCGCGCTGACGAGATCAATCCGGCG 6008
      162 HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHisCysAla 181
      6009 GAAAGAAGCCGCTCGCGCTTCAAAGCCACCAAACTGGCTTTGTCTTACTGTC--- 6065
      182 ValAlaProValGlyMet-----HisIleAlaAsnGlyMetTyrGlyLeuIleu 198
      6066 ---GCACCTCCCGAATGTTCCGTCGACATGTGATCGGACATGAACGCGCTGCATG 6122
      199 ValGluProLysGluGlyLeuPro-----LysValAspLysGlu 211
      6123 GTTCTGCGCGCGCGCGGCTGCATGACGCAAGGCAACAGTCACTTGTGACAAAGTGC 6182
      212 TyrTyrValMetGlnGlyAspPheTyrThrLys-----GlyLys 224
      6183 TATTAATGTCGCGCAGCAGCATTTCTATGTCCCGCTATAGAAACGCAATACAGACC 6242
      225 TyrGlyGluGlnGlyLeuGlnProPheAspMetLysValAlaIleArgGlu---AspAla 243
      6243 TATGAAGCGCGCGCGCATGCTATGAGATACGTCAAAGTATGCCACCTGACGCCG 6302
      244 GluTyrValValPheAsnGlySerValGlyValAlaLeuThrGlyGluAsnAlaLeuLysAla 263
      6303 ACCCATGTGTCTTCATGATGTCGTCGTCGTCCTTTCAGGATGCAAGAGCCCTGACGCCG 6362
      264 LysValGlyGluThrValArgLysPheValGlyAsnGlyGlyProAsnLeuThrSerSer 283
      6363 AAATCCGCGCAAAAGTCTGATATC-----CACTCGCAAGCCAAACCGGATACCCGT 6416
      284 PheHisValIleGlyGluIlePheAspLysValHisPheGluGlyGlyValAsn 303
      6417 CCCCACTGATCGCGCGCATGCGGATTAATGTCTGGCGACCGGCAATCAATACGCCG 6476
      304 HisAsnIle-----GlnThrThrLeuIleProAlaGlyValAlaIleThrGluPhe 321
      6477 CCAGATGTCCACCAAGAACCTGTTCATTCGCGCGCTGCGCGGCGGACGCTTGTAC 6536
      322 LysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsn 341
      6537 ACCTTCGCGCAGCGCGGATCTACGCTATGTCAATATACTGATGAAAGCTTTGAG 6596
      342 LysGlyValAlaLeuGlyIleLeuLysValGlyGlyGluLysAsnHisGlyTyrSerHis 361
      6597 CTGGCGCGCGGATGCCCACTTCAGATGACGCGGCAATGAGATGACACTTGATGACA--- 6653
      362 LysGlnThrAspAlaValTyrLeuProGlyGly 372
      6654 -----TCGATATCGCGCCCTCGCGGA 6674

```

```

RESULT 30
BX321859/c      316050 bp      DNA      linear      BCT 23-APR-2003
LOCUS      Nitrosomonas europaea ATCC 19718, complete genome; segment 4/10.
DEFINITION      BX321859 AL954747
ACCESSION      BX321859.1 GI:30138466
VERSION

```

```

KEYWORDS      complete genome.
SOURCE      Nitrosomonas europaea ATCC 19718
ORGANISM      Nitrosomonas europaea ATCC 19718
REFERENCE      Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
AUTHORS      Nitrosomonadaceae; Nitrosomonas.
TITLE      1
JOURNAL      Chain, P., Lamerdin, J., Larimer, F., Regala, M., Land, M., Hauser, L.,
PUBMED      Hooper, A., Klotz, M., Norton, J., Sayavedra-Soto, L., Arciero, D.,
AUTHORS      Holmes, N., Whitaker, M., and Arp, D.
REFERENCE      Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and
AUTHORS      Obligate Chemolithoautotroph Nitrosomonas europaea
JOURNAL      J. Bacteriol. 185 (9), 2759-2773 (2003)
PUBMED      12700255
TITLE      2 (bases 1 to 316050)
AUTHORS      Larimer, F.
JOURNAL      Larimer, F.
PUBMED      Direct Submission
AUTHORS      Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas
JOURNAL      genome consortium, the DOE Joint Genome Institute, Production
PUBMED      Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
AUTHORS      USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
JOURNAL      1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
PUBMED      larimerf@ornl.gov
REMARK      Nitrosomonas genome consortium
FEATURES      Location/Qualifiers
SOURCE      1. 316050
mol_type="genomic DNA"
/organism="Nitrosomonas europaea ATCC 19718"
/strain="ATCC 19718"
/db_xref="taxon:228410"
/complement(639..1223)
/locus_tag="NE0846"
/complement(639..1223)
/locus_tag="NE0846"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD84757.1"
/db_xref="GI:30138467"
/translation="MIRNNSFYRIDPKSPIKANPLRNPGSLNHRPLNIFLDGIV
PYFIALCFVLIAAMEWIRWYTOTSPNPLFVYMAIPCLATLIMKYIKGKEVKRKL
GLAGELAVQFLERLAQSHI FHDIPGVFNDHVHISGGIVLETGLSKPDGE
SKLYVDGNHILNKNSTALDRNPITOVANRMRRE"
order(complement(1038..1094), complement(951..1007))
/locus_tag="NE0846"
/feature=T2, probable transmembrane helices predicted by
TMHMM2.0"
/complement(1351..2301)
/locus_tag="NE0847"
/complement(1351..2301)
/locus_tag="NE0847"
/function="COG8 COG0726"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD84758.1"
/db_xref="GI:30138468"
/translation="MKILPSLPSPAGKNSKLIVYHRLVPEPDLITGSGCIAQFEG
LSVLTNNFNIPLSEAVNKLRSGLTGRACITFDGVDADNAETALPIQKHGVTATP
FLASGLDGRMNDVNLSEVRRAKGDLDNALIGNHAIALTBQRRLTNLLNK
KYLPHARASQVNDKSTIIAADLPDLNIMISQVROLHDAGMEIGGHTLTPILASID
DAAARPEIAGEKALEAIIIDTPRLPAFYANGKRGKYLTAHYVMIMIDAGEAAVSTAM
GAARDSDMFQLPRLPTPMQSEKRYVLVNRVNRKRIQTV"
/complement(2421..2873)
/locus_tag="NE0848"
/complement(2421..2873)
/locus_tag="NE0848"
/function="interPro IPR001345 COG8 COG2062"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="phosphoglycerate mutase family"

```


gene	/protein_id="CAD84759.1" /db_xref="GI:30138469" /translation="MDLILIRHAEEDGIPDATRCLKGKQAKQKARWLEPRLPKD THIVSPARIQQTIVSALTHHEETSEQVTSATPRVANTIAMPEAGETVIVGHQPL LKRILASLLKGDSESPYRKSIIWFFSSKQDERESIILRAVMTPEIL" complement (2929. .3621) /locus_tag="NE0849" complement (2929. .3621) /locus_tag="NE0849"
CDS	/locus_tag="NE0849" /function="COG61814" /codon_start=1 /evidence=not experimental /transl_table=11 /product="similar to nodulin 21" /protein_id="CAD84760.1" /db_xref="GI:30138470" /translation="MTHDNTYSHRTGMRAAVLGANDGIVSTASLIGVASAAAD DILGAGPEAAQVARELMADALDAHLDELGHERVNAKRVQAFPSAGPILGASM PLAATIAATTHIIIPVAISSLSLTAIGTFAVLTGAARVAFMGALMAFA TALVGLFGLIA" order (Complement (3496. .3561), complement (3418. .3483) , complement (3112. .3177), complement (3034. .3099) , complement (2935. .2997)) /locus_tag="NE0849" /locus_tag="NE0849" /note="5 probable transmembrane helices predicted by TMHMM2.0" complement (3637. .4311) /locus_tag="NE0850" complement (3637. .4311) /locus_tag="NE0850" /function="InterPro IPR000379:IPR003140 COG6 COG0400" /codon_start=1 /evidence=not experimental /transl_table=11 /product="Phospholipase/Cardoylesterase" /protein_id="CAD84761.1" /db_xref="GI:30138471" /translation="MPDSSFQLSAIDITTGSNPEYTLIMHGLGANDGDPVYVQALD LBEIPIRLEFPAPQOPVITNGSYIRAMVYDIORHDFVEQDEGIRSOAIVELIE LDRDRGIPDHLILAGFSQSGAMALHTGIRHPDRLAGIILSGYLPLAKTERAHTT NRITPFMAHGNDPIVRIELAHASLQQLRETYVYTMHEYPMETVCDDGLDISRW LKTILK" /locus_tag="NE0850" complement (4390. .5502) /locus_tag="NE0851" 4390. .5502 /locus_tag="NE0851" /function="InterPro IPR002931 COG6 COG1305" /codon_start=1 /evidence=not experimental /transl_table=11 /product="Transglutaminase-like superfamily" /protein_id="CAD84762.1" /db_xref="GI:30138472" /translation="MKRRSFKLGVSSSAAAAALPPAVARASLPNQRKRSQLTLYQTT LPSAGVAALMLPLPDASDTYQPTQGNNSGKATASAFYMLPGTOSQVFPALMQKEE BTAVSGSLVKTADRTVDLHSYRAPAKISIPONTORYLQSTKTHLPDDVVKQARSVT SPANHTPQOARAIYDVNNVAVDINGRQSGSDLSLRKENSLSKGGDIHALFPE VLAASAPGPARIOYGRIGDASLANKNGKSGSDISTGHCQAAEFLSGIGVAPVNPED VARWAASLSHNHDSIQLEPKLFGSEWEMVAVNREBNDLGTIKYSAISLCLPFPF GYPIAEIDRLDLSLEPSSFSYKITTSANLVGTGAKL" complement (5615. .7336) /gene="yvgQ" /locus_tag="NE0852" complement (5615. .7336) /gene="yvgQ" /locus_tag="NE0852" /locus_tag="yvgQ" /locus_tag="NE0852" /locus_tag="yvgQ" /function="InterPro IPR000660:IPR005117 COG6 COG0155"
misc_feature	/note="Signal predicted by SignalP 2.0 HMM (Signal peptide probability 1.000) with cleavage site probability 0.844 at residue 27" 4390. .5502 /locus_tag="NE0851" 4390. .5502 /locus_tag="NE0851" /function="InterPro IPR002931 COG6 COG1305" /codon_start=1 /evidence=not experimental /transl_table=11 /product="Transglutaminase-like superfamily" /protein_id="CAD84762.1" /db_xref="GI:30138472" /translation="MKRRSFKLGVSSSAAAAALPPAVARASLPNQRKRSQLTLYQTT LPSAGVAALMLPLPDASDTYQPTQGNNSGKATASAFYMLPGTOSQVFPALMQKEE BTAVSGSLVKTADRTVDLHSYRAPAKISIPONTORYLQSTKTHLPDDVVKQARSVT SPANHTPQOARAIYDVNNVAVDINGRQSGSDLSLRKENSLSKGGDIHALFPE VLAASAPGPARIOYGRIGDASLANKNGKSGSDISTGHCQAAEFLSGIGVAPVNPED VARWAASLSHNHDSIQLEPKLFGSEWEMVAVNREBNDLGTIKYSAISLCLPFPF GYPIAEIDRLDLSLEPSSFSYKITTSANLVGTGAKL" complement (5615. .7336) /gene="yvgQ" /locus_tag="NE0852" complement (5615. .7336) /gene="yvgQ" /locus_tag="NE0852" /locus_tag="yvgQ" /locus_tag="NE0852" /locus_tag="yvgQ" /function="InterPro IPR000660:IPR005117 COG6 COG0155"

[illegible]

QY 159 AlaProGlyHisSerThrPheSerPheValLeuGlnProGlyLeuTyrValTyr 178
 Db 87836 AACCGGCTGAACAAAAAATACCTTACTGATATCCCGGCTGATTTTAT 87777
 QY 179 HisCysAlaValAlaProValGlyMetHisIleAlaSerGlyMetTyrGlyLeuIleLeu 198
 Db 87776 CACTGCTGCTGTGATTCGATATTCACACATCCACGATGATGATGATGATGATC 87717
 QY 199 ValGlnProGlySerGlu-----GlyLeuProGlyValAlaSerGlyTyrValMetGln 216
 Db 87716 GTGATCCGAAAGATGCAAATGCATTCACCAAGCAGATCTGATATGCTGATTCAG 87657
 QY 217 GlyAspPheTyrThrTyrGlyTyrGlyTyrGlyGlnGlnGlnProPheAspMetGlu 236
 Db 87656 GCTGAGCATTCAGAG-----AATCTGACGACAAAAA 87624
 QY 237 LysAlaIleArgGluAspAlaGluTyrValValPheAsnGlySerValGlyAlaLeu--- 255
 Db 87623 GCCATGATGCAGAACAAATGAGATGATGCTGATTCATGAGAGGGGTTTCAAGTATGAC 87564
 QY 256 -----ThrGlyGluAsnAlaLeuLysAlaLysValGlyGluThrValArg 270
 Db 87563 CCGGTGATGATTCGTGAAGCCACCAAGCTGCTGACGCCAAACAGGTGAACGGGTGCGC 87504
 QY 271 LeuPheValGlyAsnGlyTyrProAsnLeuThrSerSerPheHisValIleGlyGluIle 290
 Db 87503 ATTACTTCGTCAACGCCGAGCCCAATGATTCATCCCTTCAACCCCATGACGGGATC 87444
 QY 291 PheAspLysValIlePheGlnGlyGly---LysGlyGluAsnHisAsnIleGlnThrThr 309
 Db 87443 TGGATGCGGTTTACCCCAAGCGCAACCCCAAAACGTCAGTATGCCCTGCAATCCAT 87384
 QY 310 LeuIleProAlaGlyGlyAlaAlaIleThrGluPheLysValAspValProGlyAspTyr 329
 Db 87383 CTGATAGAGCGGAGATGCGGCAACGCTGATTCGATTTCCGGTTGAAGCGTCCAC 87324
 QY 330 ValLeuValAlaPheHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLys 349
 Db 87323 GCGATCGTCGATCATTCGATG---CGTCATGACACATCCGTCGATGCGCGTATTATG 87267
 QY 350 ValGlnGlyGlnGluAsnHisGlu 357
 Db 87266 TTCATTAATGACGCTGATCCGAA 87243
 RESULT 31
 AFANIR
 LOCUS AFANIR 2351 bp DNA linear BCT 03-FEB-1999
 DEFINITION A.faecalis gene for nitrite reductase, complete cds.
 ACCESSION D13155
 VERSION D13155.1 GI:398121
 KEYWORDS nitrite reductase.
 SOURCE Alcaligenes faecalis
 ORGANISM Alcaligenes faecalis
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 Alcaligenaceae; Alcaligenes.
 REFERENCE 1 (bases 1 to 2351)
 AUTHORS Nishiyama, M., Suzuki, J., Kukimoto, M., Ohnuki, T., Horinouchi, S. and
 Bepko, T.
 TITLE Cloning and characterization of a nitrite reductase gene from
 Alcaligenes faecalis and its expression in Escherichia coli
 JOURNAL J. Gen. Microbiol. 139 (Pt 4), 725-733 (1993)
 MEDLINE 93294530
 PUBMED 8515232
 2 (bases 1 to 2351)
 AUTHORS Nishiyama, M.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-1992) Makoto Nishiyama, The University of Tokyo,
 Department of Biotechnology, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113,
 Japan (Tel:03-3812-2111 (ex.5126), Fax:03-3812-0544)
 COMMENT Submitted (01-SEP-1992) to DDBJ by:
 Makoto Nishiyama
 Dept. of Agricultural Chemistry
 Faculty of Agriculture

The University of Tokyo
 1-1-1 Yayoi, Bunkyo-ku
 Tokyo 113
 Japan
 Phone: 03-3812-2111 x 5126
 Fax: 03-3812-0544.
 Location/Qualifiers
 1. 2351
 /organism="Alcaligenes faecalis"
 /mol_type="genomic DNA"
 /strain="S-6"
 /db_xref="taxon:511"
 misc_binding
 /bound_moiety="putative FNR binding site"
 RBS
 737..743
 751..1881
 /gene="nir"
 751..1881
 /gene="nir"
 /codon_start=1
 /transl_table=11
 /product="nitrite reductase"
 /protein_id="BA02440.1"
 /db_xref="GI:398122"
 /translation="MAEQMSRITLACALAGALAPVLTATTSANGOGAVRKATAAE
 IALPROKVELVDPFVNAHSQVAGSPKVEETVYEEKIVIDAGTEVNAHAFNG
 TVAPLMVVDYLETLLINPTLMNIDPHATGALGGGLTEINRGETILRF
 KATKPGFVYHCAAPGMVPMHVSGNNQAIMVLPREGIDGKALTYDKIVYVGED
 FYVPRDENGKRYKEVAPGDVETVYKMTLTPTTHYVFNCAVALGDDKMTAAVGEK
 VLVHSAQNRDTRPHLIGGIDVMAVTKFNTPDVDQETWIFGGAAGAAYTFPOOP
 GIAYVYNNHILAEFLGAAAHFVTEEMDDLTMSVLAPEGT"
 BASE COUNT 540 a 661 c 663 g 487 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,48e-19 Length: 2351
 Score: 427.50 Matches: 129
 Percent Similarity: 47.00% Conservative: 51
 Best Local Similarity: 33.68% Mismatches: 158
 Query Match: 16.28% Indels: 45
 DB: 1 Gaps: 14
 US-10-088-045-2 (1-502) x AFANIR (1-2351)
 QY 15 LeuSerAlaLeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLys 34
 Db 787 TTGGCAGAGCCGCCCTGCTGGG-----GGCTTGGCGCGGT 825
 QY 35 SerSerThrValAspAla-----AlaAlaLysThrAlaAsnAlaAspAsnAla 50
 Db 826 CTGCAACACATCCCGCTGGGACAAAGCGCGCTGCGGAAGCACTGCGCAGAAATA 885
 QY 51 AlaSerGlnIleGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaPro 70
 Db 886 GCGACCACTTCACAGCCAG---AAGGTGAGACTTGTGACCTCCCTTGGTACGGCCAT 942
 QY 71 GluValProProProValAlaAspAlaSerProAlaLysValValLysMetGluThr 90
 Db 943 AGT-----CAGTTGCAAGAGCGGACCCCAAGGTGTCGAATTCACCATGTCATC 993
 QY 91 ValGlnLysValMetArgLeuAlaAsp---GlyValGluTyrGlnPheThrPheGly 109
 Db 994 GAGGAAAAAGAAATCGTCATCATGATGCGGACCAAGCAAGTTCACCGCATGGCATCAAC 1053
 QY 110 GlyGlnValProGlyGlnMetIleArgValArgGlnGlyAspThrIleGluValGlnPhe 129
 Db 1054 GGCACCGTACACAGACCGCTGATGTCGTGCATCAGACATATTCGACGACGACATC 1113
 QY 130 SerAsnHisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyPro 149
 Db 1114 ATCAACCTGAAACCAACACGCTGATGCACAAATATTCATTCATGGGACCGGGTCA 1173
 QY 150 GlyGlyGlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLys 169

```

Db      1174 TTGGGGGGGGGGGGGGGGGGTGAACCAATCAATCCGGGAGAAAAGACCAATCTCGCCCTTCAAG 1233
Qy      170  AAlaLeuGlnProGlyLeuThyValTyrHisCysAlaValAlaProValGlyMet----- 187
Db      1234 GCGACCAAGCCCGCGCTCTTCCTACCACTGC-----GCACCTCCCGGAATGGTTCCC 1287
Qy      188  ---HisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlnProGlyGlyLeuPro 206
Db      1288 TGGCATGTGTATCGGCGCATGATGATGCGATGATGCTGCTCGCGCGAGGGTCTTGCAAT 1347
Qy      207  -----LysValAspGlyGlyTyrTyrTyrValMetGlnGlyAspPhe 219
Db      1348 GACGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1407
Qy      220 TTTT-----LysGlyLeuTyrGlyGlu-----GlnGlyLeuGlnPro 232
Db      1408 TATGTATCCCGCGCGACGAGAAAGCAATACAGAAATACAGAGCGCGCGCGCGCTTAT 1467
Qy      233 PheAspMetGlnGlyValAlaIleArgGlu---AspAlaGlyTyrValAlaPheAsnGlySer 251
Db      1468 GAAGACACCGGTCAAGGTGATGCGCAGCTCTCACTCCGACCATGTGTTTCAACGGCGCT 1527
Qy      252 ValGlyAlaLeuThyGlyGlyLeuAlaLeuGlyAlaValGlyGlyLeuValAlaGlyLeu 271
Db      1528 GTGGGGCGCACTGACTGCGCAAGAGCGATGACGCGCGGTGGCGAG-----AAAGTC 1581
Qy      272 PheValGlyAsnGlyGlyProAsnLeuThySerPheHisValIleGlyGlyLeuPhe 291
Db      1582 CTGATCTCTCCACTCCGACGCGCAAGCGCGATACAGACCAACATCTGATCGGGGGCGATGG 1641
Qy      292 AspGlyValHisPheGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 309
Db      1642 GATTTATGTCTGGCGCGACCGCGCAAGTTCATACGCGCGCGCGCGCATGACGAAACTGG 1701
Qy      310 LeuIleProAlaGlyGlyAlaAlaIleThyGlnPheGlyValAspValProGlyAspTyr 329
Db      1702 TTCAATTCGGGTGGTCCCGCGCGAGCGCTTACACGTTCCAGACGCGCGCATCTAC 1761
Qy      330 ValLeuValAspHisAlaIlePheArgAlaPheAsnGlyAlaLeuGlyIleLeuGly 349
Db      1762 GCGTATGTATACCAATCTGATGAGCGTTTGAATCGCGCGCTCCGCCCACTTCAAG 1821
Qy      350 ValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 369
Db      1822 GTTCAGGGGTGAATGGAACGATCTGATGACG-----TCGGTTCTCGCA 1866
Qy      370 ProGlyGly 372
Db      1867 CCATCTGGC 1875

RESULT 32
AP005214/c 313200 bp DNA linear BCT 20-DEC-2002
LOCUS      AP005214 Corynebacterium efficiens YS-314 DNA, complete genome, section 1/11.
DEFINITION Corynebacterium efficiens YS-314 DNA, complete genome, section 1/11.
ACCESSION AP005214 BAO00035
VERSION    AP005214.1 GI:23491835
KEYWORDS
SOURCE     Corynebacterium efficiens YS-314
ORGANISM   Corynebacterium efficiens YS-314
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Corynebacterium.
REFERENCE
  1 Kawarabayasi, Y., Yamazaki, T., Hino, Y., Kikuchi, H., Nakamura, Y.,
    Iike, K., Suzuki, M., Mashima, J., Itoh, T., Yamagishi, A., Nishio, Y.,
    Ueda, Y., and Sugimoto, S.
    The entire genomic sequence of Corynebacterium efficiens YS-314
    Published Only in Database (2002)
    2 (bases 1 to 313200)
    Kawarabayasi, Y., Yamazaki, T., Hino, Y., Kikuchi, H. and
    Director-General of Biotechnology Center.
    Direct Submission
  
```

JOURNAL

Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, Nishihara 2-49-10, Shibusawa-ku, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, Tel: 81-3-3-3481-1933, Fax: 81-3-3481-8424) Kawarabayasi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8565 Japan

Nakamura, Y., Iike, K., Suzuki, M. and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan Itoh, T. is at the Japan Biological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan

Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo, 192-0392 Japan

Nishio, Y., Ueda, Y. and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8681 Japan

The other authors are at the National Institute of Technology and Evaluation, Shibusawa-ku, Tokyo, 151-0066 Japan.

FEATURES

source

CDS

```

1..313200
/organism="Corynebacterium efficiens YS-314"
/mol_type="genomic DNA"
/strain="YS-314"
/db_xref="taxon:196164"
1..1722
/note="CE0001, similar to AF222789-7|AAF33692.1| percent identity: 48 in 567 aa"
/codon_start=1
/translation="putative chromosomal replication initiator protein DnaA"
/protein_id="BAC16811.1"
/db_xref="GI:23491836"
/translation="MSQSSSLLETLRWQVNDLMSQODSGIPLPTPOAYLNTA KPIALVDYATLSTHALAKAIEHDLRESLTKTSMKGRFSALVSEPRDEDP NAPPAPRODELNLYPTPGQPOSPOGQQCQGHYPQVEVRAAPAPHQGQQAQHQ PRANAPQOYVVGESRGOASQAGMBQTHSPAPDQGPDPAPVEPPQPAHPLRI PRETPAHNPRESLNPRTYFENFVIGFENRPAANAAYAVASPAKAFNPLFISGGG LGKTHLAVNGVYAOELQPLRIKYVSESEFPNDYINSVDRDROSPKRRYNDILM VDDIOLAGKEGTQSEPRFRALHQAKQIILSSDPKRCITLEDRLRTREGGLI TDIOPLDETALIMKKAQADGTHVDSBELILSRRESSIRELEGALIRVATSSL INQPLDKEMALVALDILPEPDEMTLPAVMEVAYEYFKISVDTLRGAKYRAVAHA ROLAMYLCRELIDLSLPKICGVFGCKHITVWYADRKIROEMTERDRTDEIQOLTOI IKSRRGN"
complement(2289..2603)
/note="CE0002"
/codon_start=1
/translation="hypothetical protein"
/protein_id="BAC16812.1"
/db_xref="GI:23491837"
/translation="WMILHADIKLAEKVIKCRRTGCTGCHDPTOPDPADTVVWNGRL PWTRAIGGVVLEHLIPNVGHULTVREFVYSFLCRSEKCEARFSLFTNVSPPHPLLV KF"
2592..3776
/note="CE0003, similar to L39923-25|AAB53142.1| percent identity: 50 in 390 aa"
/codon_start=1
/translation="putative DNA polymerase III beta subunit"
/protein_id="BAC16813.1"
/db_xref="GI:23491838"
/translation="MEPNVSPFVAVREDLVTSVGVANRLPTKPTOPVIRLMTADD EGIENAGDVEYSRVRSAVAYGRJAVAKGILTEVITGTPKRPVPRIDGSKYLV SCGSASFELPSIPLDYPPPLSPERTSINPKLTETVTOVASPAAGRDSIPMLTGV MEIHGNGOISLAATDRFLRLRTFEMWDASGIEKLIIPAKTLLNARLTDTGIANE IGIIVGTQEOVREGFLGIRTDNRESTRMLDADFPNIQPLPKHETIAISVEIAPLL DAVRVSIVADRNAOIIIMFSDGOVITISAGSDAGHAEVLPICATGMDLKAFFPGY LKDGSLVPTSRVAVGTFEPSPRPMIMEPEPEFPADBRGVFGFTDTYFTYLLMVRV LK"
3878..5074
/note="CE0004, similar to AF222789-9|AAF33693.1| percent identity: 55 in 390 aa"
/codon_start=1
  
```

/transl_table=11
/product="putative DNA replication and repair protein
Ref"
/protein_id="BAC16814.1"
/db_xref="GI:23491839"
/translation="MYTRSLRLDFRMPPELKYELKGTITFIIRNGFGKTNIVEAIG
YLHLSSHRVSTDAFPAVANAADARISAVANOGRELAALLIKPHANAGQINIRKV
KSPRELKATVLFAPEDLALVRGEPAERRYLDIIRPRMAGVKADYDKYKQ
RNALKTATIALRPGYGTIEGAALALITLDTMQLALGAEVNAARFALVQDSQIR
DAYOTIAESRPAVNYKTTIDQALQFEGFDGITEATLTLEAAROREIERGMSL
VGPBHDLELHIGOPAKGFSHGEMTSPALSMRIAEFNLIRSDGDPILLIDVSE
LDAARREKVLGIARDAEQVITTAASDDELPANLADAITAHTVTVPDTDAGRISLDE
QDPDQ"

CDS
5071..5652
/note="CE0005, similar to AF222789-10|AAF33696.1| percent
identity: 39 in 188 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC16815.1"
/db_xref="GI:23491840"
/translation="MTEDDFIAQAFERVRAQRNRGRVVDLNNRDAFRRTPRIG
DRANGSGGGGVDRRGRPTGDCGRKYVRSABELGAILNRISRGRMSDIAAG
WTSHWAEIVGPKIAQHTKVKEMTKDKLFTCTSTAWATNLRMQKQILQVIAEKGP
DVIAELRIFGQPASWRKGPLHVKGRRPDTYG"
5786..7849
/note="CE0006, similar to X92503-7|CAA63253.1| percent
identity: 72 in 685 aa"
/codon_start=1
/transl_table=11
/product="putative DNA gyrase subunit B"
/protein_id="BAC16816.1"
/db_xref="GI:23491841"
/translation="MANTENYDASSTITLLEGLAVRKRPMTYIGTSGRLHLIME
VUNSDMEAGADRVDTLLLEDGVEVDNDRGIPVEMHAGCAPVQVYVMOLGANA
RGKDSBYAVSGELHGVISVNAALSTRVNAHKKCKKHYQNPNSIPEDVLEGANA
RGTGTSIRFWDALFETDPDFDTSIRLQEMAFNKGILMTLTDNRASDEBLEA
LAOGDTAVELSDQIDADETQLVDAEVDAPKPKRKKIKIPIHNGLEDEYVNL
NRTKPIHPSIVSFEAKAEVEIAEQMSYKSYHTANTINTFEGTHGEGFS
ALTSIMRWYAREHKLKEKEANLTGDDCRGLSAVSVRSEPOFEOCTKXGNEI
KSPVORANHEIGHMLEANPAKVIINKVAGSDSMFOAILPRLKTLNVERKLDVY
PKLADCRSKDEKSELYIEGDSAGGSASGSDMFOAILPRLKTLNVERKLDVY
LKNSVQAILTALGTGIDHFDINLRYHIVLMAADVGQIATLLTLFRFMD
LIAEGHYLLAQPLYLKMQRGEFGFAYSDSEBDQDEGLAAGRKINKDDGIQRYG
LGEINPELLETMTWDPSEVRIIRRVLDADAQRADELFTLGMDDVVARSPITNADV
RPLDV"

CDS
7945..8847
/note="CE0007"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC16817.1"
/db_xref="GI:23491842"
/translation="WNAETPTTVVNMPPSSRTPVQVVRAGEDGCGSGDKLVNMY
PGMGGYFRPIAELAAARGPVAAGELGCGQSSRAATRTTGWGHDSADFPQ
VRAKOGLGVPDHPMIFLAFSGGQISCLFARBEDRDLNGLITFGVSGSYRPRF
DPMGRGLGWLGLFGRVGGHGLGYMPGKILGRDPLGCGAQAHRMELRPHKNSL
DQLTGQIDVVALSRLELPVLPFRFNDRDPTASEALAEFLHNDVAFEDLPGL
GHRMAEPIYVADRFSFAGVDL"

CDS
9128..9709
/note="CE0008"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC16818.1"
/db_xref="GI:23491843"
/translation="MAEKVHVPGEHAHREHAGVNAVPSASAVERITVDDHSAQDA
WIGCTDPHVANNRISWGAIIAGAVTIALVLLIGIAAALGDDSSATVIGITVLGL
VIAFLVAGYVAGLAVAGGLHGLFATCATSLAVIFLAGWLGSVLAGFSGIAGTAD
TAACTVNTSODQGAADATSGEDVDQAQAEAE"

CDS
9994..10422
/note="CE0009, similar to X65467-3|CAA64662.1| percent
identity: 35 in 113 aa"

/codon_start=1
/transl_table=11
/product="putative transcription regulator"
/protein_id="BAC16819.1"
/db_xref="GI:23491844"
/translation="MAGVRORNPRTIGQLVYGOHMLTASRLDVNRILGRAMADPTR
SRLSLDAPGPAQALTELGRTRVNSNHCLACDCGIIVAEPGRRTREIADAH
LRAALNALDITLAVDENADCIDPQPVNCCAFEBGRQ"
10419..12371
/note="CE0010, similar to 274025-13|CAA98411.1| percent
identity: 55 in 627 aa"
/codon_start=1
/transl_table=11
/product="putative cation-transporting ATPase"
/protein_id="BAC16820.1"
/db_xref="GI:23491845"
/translation="MTSACGCEPTGCVNPDGPVMAAVGTGEHTPMWRDPTVPIV
ASGVALSLILEWVPGAGTASVFWISLIGASQFPFGAVRNILVTRGLIGLIM
TTSATGAVILGFVEEAAALAFYSIALEAGRAMDQFARGLRALIPATVAVTA
NTTRSVSEELLPGDRVLVAGRLATDGIHSGRSLDVSATIGSIPVEVGPQV
LAGSINISGVLEVEATPAAGTNSLTVVSLVEQARHEKDRADRLARLPVGVLI
LAAIVAVISLIGDPVMTIERALVIVVVAASPCALISVPTVVSATGAASRFGVIXS
GAAFELGSGIOHLADKCTLRNPTVADVVTGTAISRBAVUGMAALAEASTHPLA
AATTAATPEVPAAEVAVAGSITGLVGTIRIAGVSPMLAPGLVGTVEDEBGM
TGVVVIHQDPVGVIGVDELREPVDVIRLETQGFITMLDQRTAEALATAVE
IGDVHAQLDRPDKARAVAGIAGRPVAMIGDINDAPALASADVIGAMGKGAADAE
SADVAFTGDLRLVTPALAHARRGRTINQNIIVLSAIIIVLPLAITGALAAVVL
VHEVAEVIYIINGLRAARAGTPALS"
complement(12368..13030)
/note="CE0011"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC16821.1"

CDS

Alignment Scores:

Pred. No.:	1..156-14	Length:	313200
Score:	390.50	Matches:	108
Percent Similarity:	45.06%	Conservative:	47
Best Local Similarity:	31.40%	Mismatches:	130
Query Match:	14.87%	Indels:	59
DB:	1	Gaps:	13

US-10-088-045-2 (1-502) x AP0052214 (1-313200)

Qy	31	AlaGlnProLysSerSerThrValaAaAlaAlaAlaThrAlaAaAlaAaAla	50
Db	298126	GCCGCCCGCGCGGCGGAACCAACCGCGCGCGGTGACGTACCGGAGTC--GGCGCC	298070
Qy	51	AlaSerGlnGluHisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaPro	70
Db	298069	CGCAGCGTGACACGACGAGGCGTTCACCCCGGTCAACCGGCTCGT-----	298025
Qy	71	GluValProProProValaAspArgAspHisProAlaLysValValaLysMetGluThr	90
Db	298024	---GCCCGCGCGCGCGCGGTACCGTCCACGGA-----TACAGCTGGGATATC	297980
Qy	91	ValGluLysValaMetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlyGly	110
Db	297979	ACCGAGAGGTCCGCGACGCGGATGATGACAGCGGTGTCTGGCTTTTCAGCGG	297920
Qy	111	GlnValProGlyGlnMetIleArgValaArgGluGlyAspThrIleGluValGlnPheSer	130
Db	297919	CAGCGCCCGGACCGACCGCTCGCGGCAAGCTCGGAGCAACCTTCGCAATCACCGG--	297863
Qy	131	AsnHisProAspSerLysMetProHisAsnValaAspPheHisAlaAla---ThrGlyPro	149
Db	297862	---CAACAACAGGCGACATGATCATCTCATGATTTCCACCCCGCGAGGTCAACCCG	297806
Qy	150	GlyGlyGlyAlaGluAlaSerPheThrAlaProGlyHisIleHisSerThrPheSerPheLys	169
Db	297805	-----GACACCAACATGTGGGACACCCCGTGGGGAATCCTTGACTTACGAGTTCGTC	297752

gene
3389. .4528
/gene="pan1"
/note="synonym: VNG1187G"
3389. .4528
/gene="pan1"
/note="pan1"
/codon_start=1
/transl_table=1
/product="membrane protein"
/protein_id="AAG19561.1"
/db_xref="GI:10580721"
/translation="MTDQIGAPLGI SRDFAATAGVCTAAIGCTAPDGSEVTDTT
TAAQSGPLPTTSPEVDATGEQGNQVTLKSPAVADVPLDSMGKPVKLPRVAFAFA
EGDGSVPSP I VRTSEGODIEVTLNTDGRKPHTLHFHGSOTLAWEDDVPPTTGI R V
PEEKSTYTT PAVNPGTHLYHCYOTQRHIDMGVGI P R I D R G Y E P A D K E F M V M K V N
DSRLNMYAGDEVDSYPRTRNDPVTNAGKSPRLPHDEDSPI I I VEGDQVRLHVLN
GGMNHPHLIHNHRFRMRVEKDCQI PAAQHTMDITDAPAPERTIEFQADDPGIVL
MCHKANHVMNCTFYGGMLSGIVTKMDMSDIEGKLDPAGYDA"
4535. .5152
/gene="hcpd"
/note="synonym: VNG1188G"
4535. .5152
/gene="hcpd"
/note="hcpd"
/codon_start=1
/transl_table=1
/product="halocyanin precursor-like"
/protein_id="AAG19562.1"
/db_xref="GI:10580722"
/translation="MTDSDSAVTRRRVLQSSAGAGAAAAGIGFPAAGAAQSASIDDM
LDVNLADVDNRDETSSSEVTVGVGVANGAGVGFPAARVAVDPTVTTFEVSSTHNHV
AVESQDSDSGAGVEIINNEGYSHTSETAGVTVY FCSPRAMGMKGA VVVGADVG
GAAGSGIDMGNLGVVAGFALGIGLMPFVARALDPHNNTHE"
complement(5173) .5571
/gene="VNG1189H"
complement(5173) .5571
/gene="VNG1189H"
/note="hypothetical protein"
/codon_start=1
/transl_table=1
/product="Vng1189h"
/protein_id="AAG19563.1"
/db_xref="GI:10580723"
/translation="MTAARVNPPTVMVRALRGQGPRLRLVVGDSSTATVERHV
ALVAGAGTDADATLALRGITDGCARRALRANGVAVVAATIRNDVPTTHDILV
ATADDAGRKSVPAATTTDDSAVAVDATDT"
complement(5721) .6323
/gene="sod1"
/note="synonym: VNG1190G"
complement(5721) .6323
/gene="sod1"
/note="sod1"
/codon_start=1
/transl_table=1
/product="superoxide dismutase"
/protein_id="AAG19564.1"
/db_xref="GI:10580724"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1
/product="acyl-CoA dehydrogenase"
/protein_id="AAG19565.1"
/db_xref="GI:10580725"
/translation="MSEYELPPLPYDYALRPHISEOVLTWHDTHQGVYNGKNDAE
ETLANSREFGHASTAGALGDVTHNGSHIILHTTFWQSMSPAGDEBSGLADRIAD
FSSVNNRFAEPAASAGALLVYDSHSNTLRNVADVNDDEGALWGSHPIALDWV
EBSYVDYDGPDRGSFVDAFFEVVDWDEPTERFEQAAERFE"
5585. .7607
/gene="acd3"
/note="synonym: VNG1191G"
5585. .7607
/gene="acd3"
/note="acd3"
/codon_start=1
/transl_table=1

[illegible]

QY	151	lyglialaalaaserPhe-----Thra	159
Db	3809	--GGATACAGACCGCTTGGAGAGACGCGCTCCGACGACCGGCATCCGGCTCG	3866
QY	159	laProglyhiarhserThrPheSerPheLyalaLeuInProglyLeuTyValTYH	179
Db	3867	GCCCCGGGAGAAACACACTACCATTCGCCCAACGTCGGGAAAGCACCTTACC	3926
QY	179	isCyalaValaIaProvalaIgmehisIlealaasnlymetyTgLyLeuileuV	199
Db	3927	ACTGCCCACTAC---CAGACCCAGCGCCACATTCAGATGGGATGTACGGCATCTCCGA	3983
QY	199	alGIuProLySGluGIuLeuProLyValaIaSpluGIuTyTyTyValaImetGIuIaAsp	219
Db	3984	TGACGCCCAAG---GGGTACGAGCCCGCCGACAAGAGATTTTCATGACGCTCAAAAGCT	4040
QY	219	heTyThrTySGlyTyTyTyGIuGIuGIuInProPheAspmetGIuTyAlaI	239
Db	4041	GGACCTCGCGCGTGAACCGCAGCATGGCCGCGC-----	4072
QY	239	leargGIuAspAlaGIuTyTyVala-----ValaPheasnlySerValGIuAlaLeuThrg	257
Db	4073	-----GAGACGTCGACTACAGCCCGCGCACACCCGACGCTTACCGTCAACG	4127
QY	257	lyGIuaaAla-----LeuLySaIaIaValg	266
Db	4128	GCAAGTCGCGCGCGAACCCTCCACCCGAGACGCGCTACCGCATCGTCGAGCAGG	4187
QY	266	lyGIuThrValaIaTyLeuPheValaGIuAsnlyGIuTyProasnLeuThrSerSerPheHISV	286
Db	4188	GTCATACGCTCGCGCTCACCTGCTCACGCGCGG---TACATGAACACCCCGCTGCACA	4244
QY	286	alIleGIuIaIaPheAspLyValaHiePheGIuGIuTyTyTy-----GIuGIuA	303
Db	4245	TCCACACACACCGCTTCCGATGTGGA7AAGAGACGCGGCGGAGATCCGCGCGGCCGCC	4304
QY	303	snHisaInIleGIu---ThrThreulIeProaIaGIuAlaIaIaIeThrGIuPheI	322
Db	4305	AACACACACATGACATCACGACATGCTCCCGCCAGCGCCACACCATC---GAGTTCC	4361
QY	322	ysValaAspVala-----ProGIuAspTyValaLeu-----ValaPheHISaIaIeP	337
Db	4362	AGGCCGACCGCCGACCCCGGATCTACCTGATGCACTGCCACAAAGATCAACACGTTATGA	4421
QY	337	heargAlaPheAsnlySGIuAlaLeu	345
Db	4422	ACGCGACGTTCTACCCCGCTGCATG	4447
RESULT 34			
LOCUS	AE011334	10485 bp	DNA linear BCT 28-APR-2003
DEFINITION	Leptospira interrogans serovar lai str. 56601 chromosome I, section 143 of 397 of the complete sequence.		
ACCESSION	AE011334	AE010300	
VERSION	AE011334.1	GI:24195164	
SOURCE	Leptospira interrogans serovar lai str. 56601		
ORGANISM	Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.		
REFERENCE	1 (bases 1 to 10485)		
AUTHORS	Ren, S. X., Fu, G., Jiang, X. G., Zeng, R., Miao, Y. G., Xu, H., Zhang, Y. X., Xiong, H., Lu, G., Lu, L. F., Jiang, H. O., Jia, J., Tu, Y. F., Jiang, J. X., Gu, W. Y., Zhang, Y. Q., Cai, Z., Sheng, H. H., Yin, H. F., Zhang, Y., Zhu, G. F., Han, M., Huang, H. L., Qian, Z., Wang, S. Y., Yao, Z. J., Shen, Y., Qiang, B. Q., Xia, Q. C., Guo, X. K., Danchin, A., Saito, Giron, I., Somerville, R. L., Wen, Y. M., Shi, M. H., Chen, Z., Xu, J. G. and Zhao, G. P.		
TITLE	Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing		
JOURNAL	Nature 422 (6934), 888-893 (2003)		
REFERENCE	2 (bases 1 to 10485)		
AUTHORS	Ren, S., Fu, G., Jiang, X., Zeng, R., Miao, Y., Lu, L., Lu, G., Jiang, H., Ding, Y., Jia, J., Tu, Y., Gu, W., Cai, Z., Sheng, H., Yin, H., Zhang, Y., Zhu, G., Wang, S., Shen, Y., Qiang, B., Chen, Z., Wen, Y., Xu, J. and Zhao, G.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAR-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China		
COMMENT	Updated information will be available at our World Wide Web site (http://www.chgc.sh.cn/lep/). Comments to the authors are appreciated.		
FEATURES	Location/Qualifiers		
source	1..10485		
	/organism="Leptospira interrogans serovar lai str. 56601"		
	/mol_type="genomic DNA"		
	/strain="56601"		
	/serovar="lai"		
	/db_xref="taxon:189518"		
	/chromosome="I"		
	/note="Strain 56601 is maintained by the Institute of Epidemiology and Microbiology, Chinese Academy of Preventive Medicine"		
gene	/locus_tag="LA1552"		
CDS	complement(90..1391)		
	/locus_tag="LA1552"		
	/codon_start=1		
	/transl_table=1		
	/product="Sensory transduction histidine kinases"		
	/protein_id="AA048751.1"		
	/db_xref="GI:24195165"		
	/translation="MLPRNRVNVWIDMNLSPFKLFLSHRSYLSGVKNVIO IILIVSCRFILITIPSYLAWOGGSEVVLDTLALLVFLILLRNPRKPSYEL VSLIFILGILTKTGLGAGILMLFLVPVPCGFLERILITFMSVITLSVPSGILF IHYKMTGTSIFPQVVGANFTFLCGITFLILEINBAKQHLLTHKNKQLQ NKIPATYTESKSLISLNYKETFLKEIOHVNKHILILGLNDRKASLAKTI ESADIRIOSMGIVOEYELVINDSVKVIOTKEVLDLTNVHIFLISYRQSGISLVNITRE ALISYDKAIPCGLITNENINSFKHAPENRKIKITYSFOKSDSYTILIEADNGIQ KSEKSKMAGNSLIQFLQLRGETIETKEKNGFLVRFPF"		
gene	1635..1656		
CDS	/locus_tag="LA1553"		
	1635..1656		
	/locus_tag="LA1553"		
	/codon_start=1		
	/transl_table=1		
	/product="putative lipoprotein"		
	/protein_id="AA048752.1"		
	/db_xref="GI:24195166"		
	/translation="WKILNREITKVYVWVFWLWTIGISRGIFISCGSEKNKETSFAH VLAINDSPSPMOKIPVGATVEFINSGNPNPNAIADKNMSTKSPGNTYMSRRTKK VSPKRGVAPYFCSPFHATPDKNQVGDIVGANVPYPAKSGSKNVNVSQSTRRK VPSQYPTONADVNSPQDLVISEGIYLBERTVTTSTITIRGDRKVIIDGQFQGR NGIMNVADGVVIENTLTPRNATLNGFWYGVKRGSLTANHGDIYAFISINCV IETHSVASGSPDSGIYIGOCYPCAKIYDVVSEHNAIGSGNAGELYLIGSVKONI VGLAVNTIDRELPERPRTTILGLNVNNNPKPIAALYPSGNGILLAGLSVTSI RKQVVIENQNGIVILPNIDENFWLSHNITVDNIIVNSGADITLLGPMSTGCSBG NNEYRTLPFLKMNKGSQSWRLPVGDLSSMLGALILMOVAGSGRSPSGYKQPIP GPOLMPLGANAPVKPALTAFAEDFNILNINQVLPKRAEILIKTVPRAPASTGATILV KIGLGFPEFYMHLGLPLFAIYICWTSKSLDLKDRDLEWIKIYIVITLIVPLIS PAIYIVGGVGLALILIVTSYAKF"		
gene	3631..4011		
CDS	/locus_tag="LA1554"		
	3631..4011		
	/locus_tag="LA1554"		
	/codon_start=1		
	/transl_table=1		
	/product="hypothetical protein"		
	/protein_id="AA048753.1"		
	/db_xref="GI:24195167"		
	/translation="MLBPKRSVYKFKIFRNYRRKIMEQTVWVGPGFALLPFNYGYX PFIITLPLALASDLVREDVNSKISWTGAILLIPILGAGAYLVWAGSKIPSWL KXILYGGVGLALILIVTSYAKF"		
gene	3998..5029		
	/locus_tag="LA1555"		

gene 3998. .5029
CDS /locus_tag="LA1555"
/locus_start=1
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA048754.1"
/db_xref="GI:24195168"
/translation="MNPFRKKKSNRQFLKMGIGSAGATGAGTSSIDEDSGLLCKTIR
REIUVNKNERTSPGNPGNNYSKSWHPPEFHIDPTFLNRMELKNLEISGPILOKHS
IYEMPTLVAMNTVVAKMTFNGVLPGEPVYRAKLGQMEITLRNDSHPHSIFHPSHD
NBDGPPVAVGSEKTYQLTAPGIGFHPYHCHVPPLASHMAGLVYGLIVDPGGRPPA
HEFMFLISMDLEDJGKNDLFCMNGMAGFYDREYPIKVPVGOKVRLYIANNCEYEPVAS
FHLHAQTFPVPFRTGTLIPDHTDVVTIGQTERVILEPTLPRKGRYMHPHQTGAERK
GAMGWIVAV"
5376. .5969
gene /locus_tag="LA1556"
CDS 5376. .5969
/locus_tag="LA1556"
/locus_start=1
/codon_start=1
/transl_table=11
/product="SCOl/Senc family protein"
/protein_id="AA048755.1"
/db_xref="GI:24195169"
/translation="MKERILLISILIGILIGFGPMKAKGRSSIOEPVFEVSEQITTE
LNTNOGTGFLNKITPGKTLVYRGSHPDMDCPRLIDMSLAVKEIDGOGKTLIPVET
SYDPRKDTGLSLAKTYKQFPQERLVALTGEBSKDLSDLNAGAVSKYADAPQLEGCTTT
VHYHFIYVLDDKSRILMTFPGMDGKTLAEIRRL"
6040. .6186
gene /locus_tag="LA1557"
CDS 6040. .6186
/locus_tag="LA1557"
/locus_start=1
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA048756.1"
/db_xref="GI:24195170"
/translation="MYQPEKILHEHSIDELKNSIIDSKIVSMHFNADTNGLELFF
QQLY"
6229. .6345
gene /locus_tag="LA1558"
CDS 6229. .6345
/locus_tag="LA1558"
/locus_start=1
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA048757.1"
/db_xref="GI:24195171"
/translation="MILPEIKKKIQDRLVEIESRSVEILLATSGSRAMWF"
6314. .6529
gene /locus_tag="LA1559"
CDS 6314. .6529
/locus_tag="LA1559"
/locus_start=1
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA048758.1"
/db_xref="GI:24195172"
/translation="MNVOELGFPESIDSDCDIRFIYGHETEMFLAVLPKKMSLEFYKK
IINLVRIPEFCARDRLGVKLLIRPSKWL"
6747. .7055
gene /locus_tag="LA1560"
CDS 6747. .7055
/locus_tag="LA1560"
/locus_start=1
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA048759.1"
/db_xref="GI:24195173"
/translation="MTFAMYGHLKPEFHGMSPLPTIFLSMGIALFEYILMVPANRIGYN
BEGYSTPQLKIQEIITISVFILFASVLVKKERIKMNHAVSFLILAAVGFAYDKTHS
"
7063. .7176
gene /locus_tag="LA1561"

```

gene
7063. .7176
/locus_tag="LA1561"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA048760.1"
/db_xref="GI:24195174"
/translation="MEHISRKNVFIKEFEVSHKNGSDLIFFVNLFTCL"
7632. .7736
/locus_tag="LA1562"
7632. .7736
/locus_tag="LA1562"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA048761.1"
/db_xref="GI:24195175"
/translation="MRONLRBELTFLIKIFRLDLDLIPKGSFFPKGT"
complement(8434. .8826)
/locus_tag="LA1563"
complement(8434. .8826)
/locus_tag="LA1563"
/note="confirmed by proteomics"
/codon_start=1
/transl_table=1
/product="class II heat shock protein (HSP20)"
/protein_id="AA048762.1"
/db_xref="GI:24195176"
/translation="MTNAVLNENKSHVTSSETKKERVRLAPRVDIYSDENITVLLAD
LPGEEDVOVOLEKDLITSGTKSNKDIOGELRYSFRTGEYKRTFTLVESVEDRI
SAYYKNVNLTLTPKKRPLTKIEVRSE"
complement(8447. .9263)

```

Alignment Scores:			
Pred. No.:	1,71e-07	Length:	10485
Score:	258.00	Matches:	78
Percent Similarity:	42.96%	Conservative:	41
Best Local Similarity:	28.16%	Mismatches:	120
Query Match:	9.82%	Indels:	38
DB:	1	Gaps:	11
US-10-088-045-2 (1-502) x AE011334 (1-10485)			
QY	75	ProValaSPARgARhPiSPROAlaLyValValValmetGluTrhValGluLyVal	94
DB	4280	CCGATCTTAAACACACT-----ATTGACATGTGACAAATGCC	4318
QY	95	MeARgLeuAlaERgLyValGluTrGlnPhETrPhEgLyGlnValProGly	114
DB	4319	TTGACGGTTGGGCACATACAGTAGTCAGACGTTTAAACGTTTAAAGTTAGTTCGGT	4378
QY	115	GlnMetLeaRgValARgGluGlyARPrTlEGluValGlnPhESerAEnHLeProAsp	134
DB	4379	CGTGTGGTTGCAGACCAAACTGGGCACAAAGATGAAATCACTTTAGAAC-----GAT	4432
QY	135	SerLyMetProHISaEnValARPhEnAlaLehTrGlyProGly---GlyGlyAla	153
DB	4433	TCTGAACATCCACATTGGATTCACTTTACAGGAAGATGATGATCCGAACGAAGCGGTGG	4492
QY	154	GluAlaSerPhETrAlaProGlyAlaTrSerThrPhESerPhElyAlaLeuGlnPro	173
DB	4493	GAGCCG-----GTAGCGTTGGTAGTGAAGAAACGATATCACTGACGCGGCGCAT	4546
QY	174	GlyLeuTrValTYrHISCyValaValAlaProValGlyMetHISLeaAEnGlyMet	193
DB	4547	GGATTTCATCCGATCATCTTGTACAGTTCCCTCCATGTGCTAGTATATGCAAAAGCTTT	4606
QY	194	TyrGlyLeuLeuValGluProLyGluGlyLeuProLyValARlySGluTYr	213
DB	4607	TACGAGAGTCTGATCGGATCCACCTGAGAGTCGTCGCGCTCACATGAATTTATGCTG	4666
QY	214	ValMetGlnGly--ARPhETrThrLySGlyLyETrGlyGlnGlnGlyLeuGlnPro	232

```

Db      4667 ATTCTTTCGATGGATCTAGACAAAGGAAAG----- 4702
Qy      233 pheasmetcgluylsalalleargluaspalaaglutryvalvalpheasnglyserval 252
Db      4703 -----AATGATCTTTTGTGGAAAGGAAATGGCT 4732
Qy      253 gllyalaleuthrclgluasnaleuylsalalvalglgluthrvalargleupe 272
Db      4733 GGT---TTTACGATCGTATTCGATTAAGTTCAGTACAGTCAAAAGGACCTCTTTAT 4789
Qy      273 valgluasnclgllypro---asenuthrserSerPhehlsvalileglguileppe 291
Db      4790 ATCCGAAACATGCTGTAATATGACCCGTCCTCTTCATCATACACGCTGACATTC 4849
Qy      292 asplysalalphehclgllygllygluasnhsasnileglthrthrleulle 311
Db      4850 GACGCG-----TTTGAACCGGAGCTAATTGATTCCTGACGATCATACGACGTAGT 4903
Qy      312 proalagllyala-----alaliethrglupheylsalaspyalproglaspytyr 329
Db      4904 ACTCTGGGGCAACAGAAAGGTGATTTAGAAATTCACCTCCAAAAGAGGAGGTAT 4963
Qy      330 valleuvalasphlsalalleargluaspalaaglutryval 346
Db      4964 ATGTTTCATCTCAACAAACAAATGCG---GAGAGGGGGCTATGGGT 5011

RESULT 35
AF339049 537 bp DNA linear BCT 15-MAY-2001
LOCUS Nitrosomonas sp. TA-9211-NH4 putative dissimilatory nitrite
DEFINITION reductase (nitrK) gene, partial cds.
ACCESSION AF339049
VERSION AF339049.1 GI:14039642
KEYWORDS
SOURCE
ORGANISM Nitrosomonas sp. TA-9211-NH4
            Nitrosomonas sp. TA-9211-NH4
            Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
            Nitrosomonadaceae; Nitrosomonas.
REFERENCE 1 (bases 1 to 537)
AUTHORS Casciotti, K.L. and Ward, B.B.
TITLE Dissimilatory nitrite reductase genes from autotrophic
        ammonia-oxidizing bacteria
JOURNAL Appl. Environ. Microbiol. 67 (5), 2213-2221 (2001)
PUBMED 11319103
REFERENCES 2 (bases 1 to 537)
AUTHORS Casciotti, K.L. and Ward, B.B.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2001) Geosciences, Princeton University, Guyot
        Hall, Princeton, NJ 08544, USA
FEATURES
source location/Qualifiers
1.537
/organism="Nitrosomonas sp. TA-9211-NH4"
/mol_type="genomic DNA"
/strain="TA-9211-NH4"
/db_xref="taxon:159355"
<1..>537
/gene="nitrK"
<1..>537
/gene="nitrK"
/codon_start=2
/transl_table=11
/product="putative dissimilatory nitrite reductase"
/protein_id="AAK53363.1"
/db_xref="GI:14039643"
/translation="VYHCAPEGVPMWVVSNGHGTIMVLPRNGLKNPECKPLHYDRY
YIGENDFYIPKDENGKYTDSPGEAFADTTIMRKLIPNHVFNRGVALTGKAMT
AKGVETVMSHSEANDRTPHILIGLGDVWETGKSNPEKDLFTWFIKSGSAGAL
YTFRPGVAYVYNNHNLIA"
BASE COUNT 122 a 165 c 151 g 98 t 1 others
ORIGIN
Alignment Scores:

```

```

Pred. No.: 2,32e-08
Score: 249.00
Percent Similarity: 52.20%
Best Local Similarity: 40.11%
Query Match: 9,48%
DB: 1
Gaps: 11

US-10-088-045-2 (1-502) x AF339049 (1-537)
Qy      177 VALTYRHISCYSAALAVAlaProValGlymet-----HisileAlaenglymet 193
Db      2 GCTATCATTTGC-----GCCCGAGAGGAATGGTCCCTGGCACGCTGTTTCGGGAATG 55
Qy      194 TYRGlyleuileleuvalgluproblysgluglyleuprolys----- 207
Db      56 CATGGAGCATCATGTTCTTCGCAGAAATGGCTCAAGAAATCCGAGGGAACCGCTG 115
Qy      208 ---valasplysalalphehclgllygllygluasnhsasnileglthrthrleulle 222
Db      116 CACTACGACCGCATCTACTACATCGCGGAGAAATGATTTCTACATTCGAGAGGAGAAC 175
Qy      223 GLYLYTYR-----GLYGLUGLNGLYleuGlnProPheasp---Met 235
Db      176 GGAATATACAAACATTACGATTCCTCCGCGAAGCCTTTGCCGACACACGATATCATG 235
Qy      236 GLUYSALAlaleargluaspalaaglutryvalvalpheasnglyserValglialaleu 255
Db      236 CGCAAGCTGATC-----CCGAATCATGTCGTTTAAAGCGCGCGTGGTGGCTC 286
Qy      256 ThrGlYgluasnaleuylsalalysvalglgluthrvalargleupevalgllyasn 275
Db      287 ACCGCAAGAACGCAATGACGCGCAAGGTGGTGAACCGTG---ATGTTCCGTG---CAT 340
Qy      276 GLYGLYProaspLeuThrserSerPhehlsvalileglguilepheaspysvalhis 295
Db      341 TCCGAAGCCCAACCGGACACCGCCCGCATATCATCGCGGCTTGGCACTACGTC--- 397
Qy      296 Phegluglygllyleuglygluasn-----Hisasnileglthrthrleullepro 312
Db      398 TGGAGACCGCGCAAGTTVAGCAACCGCGGAGAAAGATCTGAGACCTCGTTATCCGG 457
Qy      313 AlaglYglYAlaAlaAlaIleThrGlupheylsalaspyalproglaspytyrvalleuval 332
Db      458 GCGGATCAGCCGAGCGGCGCTTACACCTTCGCGACCGCGGCTTACGCTACGTC 517
Qy      333 asphhis 334
Db      518 AACCAAC 523

RESULT 36
AP003594/c 343250 bp DNA linear BCT 28-NOV-2001
LOCUS Nostoc sp. PCC 7120 DNA, complete genome, section 14/19.
DEFINITION AP003594 BA000019
ACCESSION AP003594.1 GI:17132824
VERSION
KEYWORDS
SOURCE
ORGANISM Nostoc sp. PCC 7120
            Nostoc sp. PCC 7120
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE 1
AUTHORS Kaneko, T., Nakamura, Y., Molk, C.P., Kuritz, T., Sasamoto, S.,
        Watanabe, A., Iriuch, M., Ishikawa, A., Kawashima, K., Kimura, T.,
        Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,
        Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M.,
        Yasuda, M. and Tabata, S.
TITLE Complete genomic sequence of the filamentous nitrogen-fixing
        cyanobacterium Anabaena sp. strain PCC 7120
JOURNAL DNA Res. 8 (5), 205-213 (2001)
PUBMED 21595285
MEDLINE 11759840
ORIGIN 2 (bases 1 to 343250)
REFERENCE Kaneko, T.
AUTHORS Direct Submission
TITLE

```

JOURNAL

Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/cyanobase/
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
Location/Qualifiers

FEATURES

Source

1. 343250

/organism="Noctoc sp. PCC 7120"

/mol_type="genomic DNA"

/db_xref="taxon:103690"

/note="Synonym: Anabaena sp. PCC7120"

/complement (188. .6253)

/gene="a13691"

complement (188. .6253)

/gene="a13691"

/note="ORF ID: a13691"

/codon_start=1

/transl_table=11

/product="serine/threonine kinase with two-component sensor domain"

/protein_id="BAB75390.1"

/db_xref="GI:17132825"

/translation="MESVKNTYVDLAIAPGFIIEIYHSGKTVYRAVRADQOP
VHKLREYPTSELIQFRNOYAIANKIKIPIVIMLSLEPRNSYAMWEDGGIS
LRDVAQOOSLITELITITLADLHHIYQRIHKIDIKPANIINPERKQYLDF
SIASLPERONTISPIILEGTALVSPEOTGMNNGIDYRSPYSGLVFEELTGE
LPTTDDPMELVCHIAKOPDKFTIPSCORRYANRPSNGSESTKLPMNETIYVGE
IYMKAKNAEDRYQNALGKTDLEKCLAEQETGKVEYTTISRDYCDFIIPDKY
GRABEITLLAENRVSIGATEMLVAGFSGIKETAIVNEKPIVAKRQYFIOGKED
QFORNIPEFAVOAFNLMGOLTESPTQIQWRSKILEVGRKGVILEVLELEI
IGKOPAPKSLGSLQNRPNILPKFKFTQVTSAPHSPIFLDDLOMADSGSLKLIOL
MNDGTYLIGAYRNEVNPAPHMLIVNEIKGNNGITIKIDLAFLNOLVNTLVAT
LKSEETASOIALIVEKTOGNBPFTTQPKAHONLIQPNYDLGCMQCDIAQVNHQ
ALTDVYAFASQKLPRAQTQEVLOLAATGMOFDELTAIYESSPMKTAQWKA
LQEGLIIPQSEVKFVGEQEIIDHAQTEIIVYKFLHDRVQAAYCLIPRARALAH
HHIGQLLOKISPAAREHEIETVNOANGITITITQOERDELAOLNNAQARAT
AYQSRVADTSMITLGETAMORQOYLTILHELAELIALCGDEVQMOLITVNH
AKQSLRVOYQVOKIOLTSRNELEIATGSKFLOELGSLPDHPPEVQOAREEI
NRLIGRLGEPINLPMTDPEKLAIMQISSIIPACVMGSLIYLIVLOKLSLO
FNGSLRSAGVSYAHPHLSSTTQNMALAOQLGOTIYQLASPEPKAIRATFIVLGSY
FNGTSHLODTLPITQEGCTAGLETGNTLEFGNVQFNGNAYWSGESLSEFSQCT
YHQLDLFNLVTSAMHCFYVWESALILIGSABEISLRQDAVEKIVTEALANDLR
LFOQLYLRVLNLADPRAKQDAERKQYLTGCMGSAEPIFYFDSLTALAEHQ
PAGRSOMQVRENOAILDKADHAPMNYLHKQVLAETFRLLQETEMEYEMAI
KKAENKYLDELANELAKFYLDMGKEKVAIVVQEAAYCAWPEAKKATNDLEK
YPOLLOPILOERPKRPFQETIAVPGSSSTLSSTGTSISITPLDPASTLKAQVYS
TSLEBELITNLTEILNENSGAKSALLPQEDIWQIKAWTLSNFOPNSSSESTQYLE
SOPLTECEDIPKNIYYVNTQQTIVIDNQTDIPGILGMYLQHPQOSVFCPPMNO
GHLVGLIYLENRLTRGVFTSDLEVRILLSAQAVLEKARLYQESQTKAQOLKSLK
OOKILFNVNQMSQDLNAILFCVTQNTIRILIDVDRVGIYOPHLQVNEGEFVAD
VSPAPSAVALAKVQDHQCEGNVANIYKQGIICAITVQOSEILDKROLIAQPHVAS
LVYPIHQBELKGLICIHOCDRPOMEPMORAOVYGOMGIALKOTDLITQTOQA
TOLEHTLOHLQOTQLOLVQNEKRSALGNLAVGAHEMNNPLGFIIVSLQTOPTQDA
MEHLKYQESLPHRBDKILHAAQIIDLYLEBPLKLEAMVACDLKPKNISLSTLTF
SRADKYKIPFNHKGIDSTILLKRLKANDRPAIEVTEYGNLPIECFPGOALN
VFNNILANAIADLEDANGLSVAEITANNRRIITRTQVNNHVRISITDGMISSEL
KOHLEHLFTTKGEKGTGLGALAIARQIVVERKGALEVNSQLGQGTETPIITPAE
SANKY"

gene

6804. .7658

CDS

/gene="a13692"

6804. .7658

/gene="a13692"

/note="ORF ID: a13692"

hypothetical protein"

/codon_start=1

/transl_table=11

/protein_id="BAB75391.1"

/db_xref="GI:17132826"

/translation="MNPPOIGVALALALTFQFKVUGATLMTVAQKILFPAKYR
GLRTQNIETPLIYSLNIVSVTRIVIVAILDFGGLTSPALALAAVAGALGAAG
GLIANPAGAFILIFRFKVGDFITNAGTIGVSEGLFTTDTITTPDNVITIAANKI
PADNIQFSAHPFRVLDLQALHHDVDHQAIALILKARISQIPNVIEENPAPDVEILTF

gene

NLAPVLAVRPYCNDHWQVYFDTNKAIREFTGEAGYPIPEQRYGFSVPENGTSSV
ISSALS"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

CDS

/gene="a13693"

[illegible]

gene
CDS
10496..11362
/locus_tag="BC1582"
10496..11362
ATVSTVKNVTYGGIILHTCCRS"
LVNEVALTGNENAIKPKSCYHLERRKRFILPKRMQDNPBLELVNCFQSHIVSGNA
QBEITIVTSGSSMTKAQKRIQVYGNNEITSKALLAEYMKIGMIPEFVSQRPVDE
FAYPTTRRGSGSSMTKAQKRIQVYGNNEITSKALLAEYMKIGMIPEFVSQRPVDE

Alignment Scores:

Pred. No.:	6.81e-05	Length:	347456
Score:	245.00	Matches:	108
Percent Similarity:	35.26%	Conservative:	57
Best Local Similarity:	23.08%	Mismatches:	159
Query Match:	9.33%	Indels:	144
DB:	1	Gaps:	19

QY	67	ThrHisAlaProGluValProProValAspArgAspHisProAlaIysValValVal	186
DB	268075	ACTGAAACGCGTACACACCCATTGGAAAGTTGAAAAAGACCCAGAACGTTACTTTAATACG	26813
QY	87	LysMetGluThrValGluIuIysValMetArgLeuAlaAspGlyValGluIuIysPheTyr	106
DB	268135	AAA-----GAAAGAAAGCAAAATTAAGTAAACGGTGTATTGTTCCAGTCTCG	26818
QY	107	ThrPheGlyGlyGlnValProGlyGlnMetIleArgValAlaArgIuIuIysAspThrIleGlu	126
DB	268183	ACATTTAATGGCTCATCTCTCGTGTACAGAAATTCGGGTGAAAAAGGTGAAAGGTGAAA	26824
QY	127	ValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPheHisAlaAla	146
DB	268243	GTGACATTAAATAAT-----GAATTATCTGCACCCAGTATCTATCTATTGCGCATGATAT	26829
QY	147	ThrGlyProGlyGly-----GlyAlaGluAlaSerPheThrAlaProGly	161
DB	268297	CCTGTCCCAAAATACATGATGGAATTCACAGCGGTGACACAAATGACAGTTGAACCCAGA	26835
QY	162	HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuIuIysValTyrHisCysAla	181
DB	268357	AAAGCTTTCACCTACCAATTTGAAGGCAACCTACCCAGGAACGATCGTGATTCCTCGCAT	26841
QY	182	ValAlaProValGlyMetHisIleAlaAsnGlyMetCysGlyLeuIleLeuValGluPro	201
DB	268417	CAAGATTCTGTAAAT---CAATTACGATAGAGGTTTGTTGATGAGACCGCTCATTTGAGAAAT	26847
QY	202	LysGluGlyLeuProLysValAspLysGlyTyrTyrValMetGlnIysAspPhe-----	219
DB	268474	ACAAAG-----GAAAAATATGATTAAGATTACACATTTAATGTTGATGATGAGGTTACA	26852
QY	220	-----TyrThrLysGlyLysTyrGlyGlu--	227
DB	268528	GATTAAGAGAAATGATATAGACGTTAAAAAGAAATGACAAAGGCAAAATAGGTAATAAA	26858
QY	227	-----	227
DB	268588	TCTAAAGGTAATGAAAAATGGCGAAAGAAAGATGATGATTAAGAACCGCATGATCTTCGAT	26864
QY	228	-----GlnGlyLeuGlnProPheAsp	234
DB	268648	ATGAACATGGCGCATGATAAAAAGACCTAGTAATATGAGAAAGAAATGGACCATGGAAT	26870
QY	235	MetGluIysAlaIleArgGluAspAlaGluTyrValValPhe-----AsnGlySerVal	252
DB	268708	ATGAAGATGGAAGGTCATGATGATGAGATGATGATGATTTATTCACAAATCAATGGAAGAAAGC	26876
QY	253	GlyAlaLeuThrGlyGluAsnAlaLeuLysAlaIysValAlaGlyIuIuIysThrValArgLeuPhe	272
DB	268768	GGTGAATTATGTAAGCG-----CATTTAAAGTGAATTAAGGGAATTAACGTTCTGCTTGA	26882
QY	273	ValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlu-----	289

[illegible]

CDS

<1..>537
 /gene="nitrK"
 /codon_start=2
 /transl_table=11
 /product="putative dissimilatory nitrite reductase"

BASE COUNT 106 a 176 c 166 g 89 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Beet Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
177	3,17e-07	232.00	52.78%	8.83%	537	69	26	57	28	10

US-10-088-045-2 (1-502) x AY078254 (1-537)

QY 177 ValTyrHisCysAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMet 193
 |||||
 2 GTCTATCATTTGC-----GCGCGGAGGAGCATGATCCCTGGACACGTCGTCGCGGCATG 55
 |||||
 QY 194 TyrGlyLeuIleuValGluProlyGluGlyLeuPro----- 206
 |||||
 56 CACGGCGCATATGATGCTGCGCGGAGGAGCATGCTCAAGAACGAGAGGCGCACTGTC 115
 |||||
 QY 207 LysValAspLysGluTyrTyrValMetGlnGlyAspPheTyrThrLysGlyLysTyrGly 226
 |||||
 116 AAGTCCGACACCGCTCTACTATGCGGATACGAGATTTCTACATCCG---CGGACGAG 172
 |||||
 QY 227 GluGlnGlyLeuGlnProPheAsp--MetGluLysAlaIleArgGluAspAlaGly-- 244
 |||||
 173 GACGGCAACTTCAAGAGTTCGACTCCCTGGCGAGGCTACGAGAGCATGAGAGTGC 232
 |||||
 QY 245 -----TyrValValPheAsnGlySerValGlyAlaLeuThrGly 257
 |||||
 233 ATGCGCGGCTGATCCGACGCTGATGCTTCAACGCGCGGCGGCTGACGAGGC 292
 |||||
 QY 258 GluAsnAlaLeuLysAlaLysGlyGluThrValArgLeuPheValGlyAsnGlyGly 277
 |||||
 293 GACACCGCATGAGTTCAGAGTTCGCGGAGCGGTC---CTGTTCCGTC---CATTCACG 346
 |||||
 QY 278 ProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGlu 297
 |||||
 347 GCCAACCGGACCTCCAGCCGACCTGATCGCGGCGCATGCGCATGCTGCTC---TGGGA 403
 |||||
 QY 298 GlyGlyLysGlyGluAsn-----HisAsnIleGlnThrThrLeuIleProAlaGly 314
 |||||
 404 GAGCGCAAGTTCGCGCAACCGCGCGCAAGAGTTCGAGACCTGTTTCATCCGTCGCGGT 463
 |||||
 QY 315 GlyAlaIleAlaIleThrGluPheLysValAspValProGlyAspTyrValLeuValAspHis 334
 |||||
 464 TCCGCGGCGCGGCGGCTCTACACTTCTCGACGCGGCGGCTTACGCTTACGCTACACAT 523
 |||||

RESULT 39
 AY078247 537 bp DNA linear BCT 03-MAR-2002
 LOCUS Ensifer sp. 2FB8 putative dissimilatory nitrite reductase (nitrK)
 DEFINITION
 Gene, partial cds.
 ACCESSION AY078247 GI:19072860
 VERSION AY078247.1
 KEYWORDS
 SOURCE Ensifer sp. 2FB8
 ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae; Sinorhizobium/Ensifer group; Ensifer.
 REFERENCE 1 (bases 1 to 537)

AUTHORS Song, B. and Ward, B.B.
 TITLE Nitrite reductase genes in halobenzoate degrading denitrifying bacteria
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 537)
 AUTHORS Song, B. and Ward, B.B.
 TITLE Direct Submission
 JOURNAL Submitted (08-FEB-2002) Geosciences, Princeton University, Guyot Hall, Princeton, NJ 08544, USA
 FEATURES
 source
 1..537
 /organism="Ensifer sp. 2FB8"
 /mol_type="genomic DNA"
 /strain="2FB8"
 /db_xref="taxon:126147"
 <1..>537
 /gene="nitrK"
 /codon_start=2
 /transl_table=11
 /product="putative dissimilatory nitrite reductase"
 /protein_id="AAL82506.1"
 /db_xref="GI:19072861"
 /translation="VYHCAPEGMT PMHVSQMGHTIWLPRDGLKNEKELVKSPTVY YIGNEFYI PRDEGDKFKFDSGESEETMEVNRGLI PPHVNGRVLGTGDNAMK SKYGETVL FVHSOANRSDRPHLIGHGDIYVMAIGKFRNPDDQETWLIPGRTAAGF YTFROGVAVYVNHNLIEA"

BASE COUNT 106 a 174 c 170 g 87 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Beet Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
177	5.03e-07	229.00	50.84%	8.72%	537	70	21	62	26	8

US-10-088-045-2 (1-502) x AY078247 (1-537)

QY 177 ValTyrHisCysAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMet 193
 |||||
 2 GTCTATCATTTGC-----GACCGGAGGAGCATGCTGCTGGACATCACTCGGCGCATG 55
 |||||
 QY 194 TyrGlyLeuIleuValGluProlyGluGlyLeuPro----- 206
 |||||
 56 AACGGCGCATATGATGCTGCGCGGAGGAGCATGAGAGAGAGGCGGCGGCTG 115
 |||||
 QY 207 LysValAspLysGluTyrTyrValMetGlnGlyAspPheTyrThrLys----- 222
 |||||
 116 ACCTAGACACATCTACTATGTCGCGGAGGAGCATTTCTACTCGCGCAAGACGAGGCC 175
 |||||
 QY 223 -----GlyLysTyrGlyGlnGlyLeuGlnProPheAspMetGluLysAlaIle 239
 |||||
 176 GGGAACTACAGAGAGTCAAAACCCCGGCGAAGCTTATGAAGATCTGTCAAGCGGATG 235
 |||||
 QY 240 ArgGlu---AspAlaGluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGlu 258
 |||||
 236 CGCAGCGTACCCCGGACCATGCTTCAACGAGCGGCGGCGGCTGACCGGCGAGC 295
 |||||
 QY 259 AsnAlaLeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyPro 278
 |||||
 296 CAGCTTTACACGCGGCGGCGGCGGAG-----CGTGTCTGCTGCTCATTCGACGAGCC 349
 |||||
 QY 279 AsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGly 298
 |||||
 350 AACCGCGATACCGGCGGCGGCGGAGCATGAGCGGCGATGACTATGTC---TGGCGGACC 406
 |||||
 QY 299 GlyLysGlyGluAsnHis-----AsnIleGlnThrThrLeuIleProAlaGlyGly 315
 |||||
 407 GGCAGTTCGCAACCCCGGATCTGACACGAGAAACCTGCTATTCGCGGCGGAGACC 466
 |||||


```

QY      316 AlaAlaIleThrgluPheIysValAspValProGlyAspTyrValLeuValAspHis 334
DB      467 GCGGGGCTGCTTCTACACCTTCCGCCGCGGTGTGTACCGCTACGTCACACAT 523
RESULT 40
LOCUS   AY078248
DEFINITION
Ensilifer sp. 4FB6 putative dissimilatory nitrite reductase (nirK)
gene, partial cds.
ACCESSION
AY078248
VERSION
AY078248.1 GI:19072862
KEYWORDS
Ensilifer sp. 4FB6
SOURCE
Ensilifer sp. 4FB6
ORGANISM
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensilifer group; Ensifer.
REFERENCE
1 (bases 1 to 537)
Song, B. and Ward, B.B.
Nitrite reductase genes in halobenzoate degrading denitrifying
bacteria
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 537)
Song, B. and Ward, B.B.
Direct Submision
Submitted (08-FEB-2002) Geosciences, Princeton University, Guyot
Hall, Princeton, NJ 08544, USA
LOCATION/Qualifiers
1..537
/organism="Ensilifer sp. 4FB6"
/mol_type="genomic DNA"
/strain="4FB6"
/db_xref="taxon:126148"
<1..>537
<1..>537
<1..>537
/gene="nirK"
/gene="nirK"
/codon_start=2
/transl_table=11
/product="putative dissimilatory nitrite reductase"
/protein_id="AA182507.1"
/db_xref="GI:19072863"
/translation="VYHCAPEGVVPMHVTSGMNGAIVLPRDLKDEKQPLTYDKIY
YVGEQDFVYKQDAGNYKKVETPEBAVEDVAKMRTLTPHIVFNGVVALTGDHALT
AAVGERVLVHSGANDRTPHLIGHGDVYGAIGKPRNPDLQETWLLPGTAGAAT
YTFRQGVVAVVHNLIFA"
BASE COUNT      106 a      175 c      170 g      86 t
ORIGIN
Alignment Scores:
Pred. No.:      5.86e-07      Length:      537
Score:          228.00      Matches:      70
Percent Similarity: 49.17%      Conservative: 19
Best Local Similarity: 38.67%      Mismatches: 62
Query Match:      8.68%      Indels:      30
DB:              1      Gaps:      8
US-10-088-045-2 (1-502) x AY078248 (1-537)
QY      177 ValTyrHisCysAlaValAlaProValGlyMet-----HisIleAlaAsnGlyMet 193
DB      2 GTCATCATTTGC-----GCGCCGAGGCGCATGTGCTCGACACGTCACTCGGGCATG 55
QY      194 TyrGlyLeuIleLeuValGluProLysGluGlyLeuPro----- 206
DB      56 AACGGGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 115
QY      207 LysValAspLysGlyTyrValValMetGlnGlyAspPheTyrThrLys----- 222
DB      116 ACGTAGACAAGATCTACTATGTGCGGACAGACACTTCACTGCGGACGAGAGAGGCC 175
QY      223 -----GlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMetGlyLysAlaIle 239
DB      176 GGGAACTACAAGAACTAGCAAAACCCCGGCAAGCCTATGAAGATGCTGTCAAGGCGATG 235

```

```

QY      240 ArgGlu---AspAlaGluTyrValAlaPheAsnGlySerValGlyAlaLeuThrglyGlu 258
DB      236 CCGACGCTGACCCCGACCCGACATCGTCTTCAACGGTGTGTGCGCGCTGACCGGCGAC 295
QY      259 AsnAlaLeuLysAlaLysValGlyGluThrValArgLeuPheValGlyAsnGlyPro 278
DB      296 CATGCTTGACTGCGCGCGCTGCGCGAG-----CGTGTCTGCTGCTGCTGCTGCTGCTG 349
QY      279 AsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGluGly 298
DB      350 AACCGGATACGCGACCGCATGATCGCGGCGCATGTGACTATGTC-----GGG 400
QY      299 GlyLysGlyGluAsnHisAsn-----IleGlnThrThrLeuIleProAla 313
DB      401 GCGACCGGAAATGCCGCAACCGCGGATCTGCACACAGAAACCTGCTCATTCGGGGC 460
QY      314 GlyGlyAlaAlaIleThrgluPheLysValAspValProGlyAspTyrValLeuValAsp 333
DB      461 GGAAACCGCGCGCTGCTTCTACACCTTCCGCCGCGGTGTGTACCGCTACGTCAC 520
QY      334 His 334
DB      521 CAC 523

```

Search completed: September 5, 2003, 12:57:48
Job time : 7408 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:27:06 ; Search time 42 Seconds

(without alignments)
1149.445 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626

Sequence: 1 MSKPTLIKTTLLICALSALML.....NKGQSLADVDVAKAKTKTPN 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR 76:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1388	52.9	390	2 E81062	nitrite reductase
2	1386	52.8	386	2 B81816	nitrite reductase
3	1386	52.8	392	2 A49208	nitrite reductase
4	461.5	17.6	378	2 UC4648	nitrite reductase
5	453	17.3	379	2 A48936	nitrite reductase
6	451	17.2	360	2 JG0170	nitrite reductase
7	451	17.2	363	2 S32112	nitrite reductase
8	448	17.1	336	2 JEO215	nitrite reductase
9	442	16.8	411	2 AB3095	nitrite reductase
10	442	16.8	455	2 H98191	nitrite reductase
11	439.5	16.7	376	2 A95347	nitrite reductase
12	431	16.4	376	2 AC3633	nitrite reductase
13	427.5	16.3	376	2 J39582	nitrite reductase
14	250.5	9.5	379	2 E84274	membrane protein
15	248.5	9.5	338	2 AG2298	hypothetical prote
16	222	8.5	163	1 B81167	cytochrome c552 NM
17	222	8.5	163	1 F81938	cytochrome c552 NM
18	203.5	7.7	605	2 S52253	copper resistance
19	194	7.4	500	2 B83910	hypothetical prote
20	187	7.1	449	2 A98239	hypothetical prote
21	187	7.1	449	2 AC3047	multicopper oxidas
22	174.5	6.6	449	2 C95332	probable copper-co
23	170	6.5	552	2 A51027	L-ascorbate oxidas
24	167	6.4	130	2 F87384	cytochrome c famil
25	164.5	6.3	609	1 KSPSCY	copper resistance
26	163	6.2	570	2 H87358	copper-binding pro
27	163	6.2	579	2 S11027	L-ascorbate oxidas
28	161	6.2	578	2 S66353	L-ascorbate oxidas
29	161	6.1	131	1 CCTW5T	cytochrome c552 [v

30	159.5	6.1	611	2 C82845	copper resistance
31	158.5	6.0	504	2 F70813	hypothetical prote
32	154	5.9	343	2 T35030	probable copper ox
33	153.5	5.8	587	1 KSKVAO	L-ascorbate oxidas
34	153	5.8	439	2 A83363	probable cytochrom
35	151.5	5.8	380	2 T04343	L-ascorbate oxidas
36	150	5.7	374	2 F83631	cytochrome c oxida
37	150	5.7	675	2 E83075	membrane-bound alc
38	149	5.7	632	2 F83387	copper resistance
39	146	5.6	468	1 S14271	membrane-bound alc
40	145.5	5.5	472	1 B49340	pollen-specific pr
41	145	5.5	555	2 S23763	L-ascorbate oxidas
42	144.5	5.5	567	2 T44928	L-ascorbate oxidas
43	144	5.5	582	2 T05020	coagulation factor
44	143.5	5.5	2211	1 KFB05	copA homolog - Xan
45	142	5.4	635	2 A36868	

ALIGNMENTS

RESULT 1

E81062
Nitrite reductase (EC 1.7.99.3) [NM01623 (similarity)] - Neisseria meningitidis (strain M
N/Alternate names: outer membrane protein Panl homolog
C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #ext_change 19-Jan-2001
C/Accession: E81062
R/Tetrahymena, Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.
McKee, E.K.; Hart, D.H.; Gilman, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Piazza, M.
Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappapoli, R.; V
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A/Reference number: A81000-0-0-M01D:2015755; PMID:10710307
A/Accession: E81062
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-390 <TET>
A/Cross-references: GB:AE002512; GB:AE002098; NID:97226866; PID:NAF41975.1; PID:972268
A/Experimental source: serogroup B, strain MC58
C/Genetics:
A/Genes: NMB1623
C/Keywords: oxidoreductase

Query Match 52.9%; Score 1388; DB 2; Length 390;

Best Local Similarity 70.0%; Pred. No. 5e-87;

Matches 266; Conservative 40; Mismatches 66; Indels 8; Gaps 3;

QY	11	LICALSLMLSGCSNQADKAOPKSTVDAAAKTA--NADNAASQEHOGELPVIDAIVTHA	69
DB	9	MISLFLALACG-----GEPAAQAPAEPTAPAAAEASAAQTAETPGELPVIDAIVTHA	64
QY	70	PEVPPVDRDHPAKVVKVMTVEKVKMLADVEYQFTFGQVPGQMRVREGDTIVQF	129
DB	65	PEVPPVDRDHPAKVVKVMTVEKVKMLADVEYQFTFGQVPGQMRVREGDTIVQF	124
QY	130	SNHPDSKMPNNDVPHATGCGGAASFTAPGHTSTSFALOGGLVYVCAVPMHI	189
DB	125	SNHPDSKMPNNDVPHATGCGGAASFTAPGHTSTSFALOGGLVYVCAVPMHI	184
QY	190	ANGVYGLIYVPEKGLPKVYKXYVMQGDYTKGKYEGOGLOPPDMKAKREDVEYVFN	249
DB	185	ANGVYGLIYVPEKGLPKVYKXYVMQGDYTKGKYEGOGLOPPDMKAKREDVEYVFN	244
QY	250	GSVYALTGENALAKAVGETVRLFPVNGGPNLTSFVHIGEIFDKVHEGKGGENHNIQTT	309
DB	245	GHWGALGIDNALAKAVGETVRLFPVNGGPNLTSFVHIGEIFDKVHEGKGGENHNIQTT	304
QY	310	LIRAGGALTEPFKVDVGVVYVDHAIFFAFNNGALGLVVEEENHETISHTQDAVYL	369
DB	305	IVPAGGSAIVEFVDIPGSYTLVDHSTIFRAFNKGALGLVVEEENHETISHTQDAVYL	364

```

Qy      370 PEGAPQAI DTQ EAPKTPAPA 389
          || | |||
Db      365 GNGAPAP---ASAPAPASAPA 381

```

RESULT

nitrite reductase (EC 1.9.9.3) NMA1887 [similarity] - Neisseria meningitidis
C|Species: Neisseria meningitidis
C|Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C|Accession: B01886
R|Pithall, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Moreland,
S.; Holtrop, S.; Jasele, K.; Lachter, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Naiture 400, 502-506, 2000
A|Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis*
Reference number: B01775; MDID: 20222556; PMID: 10761919

Query Match	52.8%	Score 1386	DB 2	Length 386
Best Local Similarity	68.7%	Pred. No. 6.8e-87		
Matches 263	Conservative 45	Mismatches 69	Indels 6	Gaps 2

```

QY      7  IKTLICLSLMLSGCNSOADKAAOPKSSVVDAAKTAANDNAISOHOELVIDIAIV 66
       1  MKRQALAAIISMFLAAACGGEPAAOQTPAASABAASSA---GTHAETPAGELVIDIAVT 57
QY      67  THAEVPPVPVDRDHPKAKVVMKETEYKVMRLADVEIOFTWFGQVPGQIMIRREGDTIE 126
       58  THAEVPPPAIRDYPAKVRVMETVEYKTKMKDDGEYVYMTFDDVDVPRMIRVEGDTVE 117
QY      127  VQFENHPRSKMPHNVDFHAATGPGGGAASHTABGHSTSEFKALQGLYYHCAVAPVG 186
       118  VESFNPSSTVPHAVDFHAATGQGGGAATTTAGRSSTSEFKALQGLYYHCAVAPVG 177
QY      187  MHIANGMGILLVEPKEGLPKYDKERYVYVMOQDFYTKGKYSGQGLQPFDMEKAIREDABYV 246
       178  MHIANGMGILLVEPKEGLPKYDKEFYTVQGDFTTKGKKGAQGLQPFDMEKAIREDABYV 237
QY      247  VFNGSVGALTEGNALKAKVGETVALFVNGGNGPNLTSSPHVIGEIFDKVHFEKGXGENHNI 306
       238  VFNGHVGAIAADNMLKAKAGETVMRYVNGGPNLTSSPHVIGEIFDKVYVEGKGLINENV 297
QY      307  QTLIPAGGAITEKXUDVPGDYVLVDHAIFRAIRNKAGLILKKEGEBNHEIYSHKOTDA 366
       298  QSTIVPAGSGAIVEKVDIPGSYTLVDHSIFRAIRFKAGLQGLKYEGAGENPEIMTOKLSDT 357
QY      367  VYLPBEGAQOAI DTQEA PKTPAPA 389
       358  AYAGNGAAPA---ASAPASAPA 377

```

RESULT 3

Nitrite reductase (EC 1.7.99.3) *Paeni* (similarity) - *Neisseria gonorrhoeae*
N Alternate names: outer membrane protein *Paeni*
C / Species: *Neisseria gonorrhoeae*
C / Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 15-Sep-2000
C / Accession: A49208
R / Hoehn, G.T., Clark, V.L.
I Infect. Immun. 60, 4704-4708, 1992
A / Title: The major anaerobically induced outer membrane protein of *Neisseria gonorrhoeae*
R / Reference number: A49208; MUID:93014188; PMID:1398981

A:Accession: A49208
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-392 <HOE>
A:Note: sequence extracted from NCBI backbone (NCBIP:116466)
C:Keywords: oxidoreductase

Query Match	52.8%	Score 1386	DB 2	Length 392
Best Local Similarity	69.5%	Pred. No. 6.9e-87		
Matches 264	Conservative 44	Mismatches 66	Indels 6	Gaps 3

QY 1 LICALSALMLSCSCSNQAPKAQPKSSTYDAAKTA-NADNASEHGEIPLVIDAIVTHA 69
 Db 9 MLASLFPALLAACG-----GGQAQAPEAFETPAASAEEAASSAAQAQTATTPAGELPVIDAVYTHA 64
 QY 70 PEVPPVDRDHPDAKVVVKMETVEKVMRLADGEVQFWTFGGQVPGQMIIRVREGDTIEVQF 129
 Db 65 PEVPPAIDBDYPAKVRVKMETVEKTKMKDDCEVERYWTFPDGDVPGRMIRVREGDTIEVEF 124
 QY 130 SNHPDSKMHNDHFHAATPGGGAASSTAAGHSTSPKALQPGLVNHCAYAPVCVMT 189
 Db 125 SNNSPSTVHNVDHFHAATGGGGAATPTAAGRTSTSPKALQGLVITYCAVAAPGMMH 184
 QY 190 ANGMVYLIVPEKEGIPKVDEKYVVQGDFFYTKKCYSEBQGLQPFDEMEKALREDAEYVFN 249
 Db 185 ANGMVYLIVPEKEGLPKVDEKEFYI VQGDFTTKGKKAQGLQPFDMDKA VAEQPEYVFN 244
 QY 250 GSYVALTGENALKAKVGETVRLFVNGGGPNLTSSPHVIGEIFDKVHEEGGKGENHNIQTT 309
 Db 245 GHVGSIAQDNALKAKAGETVRRVYVNGGPNLVSSPHVIGEIFDKVYVEGGKLTINENQST 304
 QY 310 LIPAGAAATTEKRVDPGDYLVLDHAIIRANNGKALGLIKKEGSENHETYSKQTDVYL 369
 Db 305 IVPAGSAAIVEKEKVIIPGSYTLVHSHIRAFNKKALQOLKVEGANENPEIMTOKLSDTAYA 364
 QY 370 PEGAPQAITQGEAPKTPAP 389
 Db 365 GSGAASA-PAASAPPAASAPA 383

RESULT 4

nitrite reductase (EC 1.7.99.3) precursor - *Achromobacter cycloclastes*
C_Species: *Achromobacter cycloclastes*
C_Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 08-Oct-1999

Ri;Chen, J.Y.; Huang, W.C.; Chang, T.; Chang, W.C.; Liu, M.Y.; Payne, W.J.; Legall, J. Biochem. Biophys. Res. Commun. 219, 423-428, 1996
A;Ittem: Cloning, characterization, and expression of the nitric oxide-generating nitric A;Reference number: JC4648; PMID:96193667; PMID:8605003

A:Molecule type: DNA
A:Residues: 1-378 <CHE>
A:Cross-references: EMBL:Z48635, NID:g1125638, PIDN:CA88564.1, PID:e140933, PID:g1125638
A:Experimental source: TAMI013
A:Note: The authors translated the codon GGT for residue 259 as His
Rifenderson, F.F., Kumar, S., Adams, E.T., Liu, M.Y., Payne, W.J., Le Gall, J.
Biochemistry 30, 7180-7185, 1991
A:Title: Amino acid sequence of nitrite reductase: a copper protein from *Achromobacter*
A:Reference number: A37260, MUID:91308101, PMID:1830217

A/Status: preliminary
A/Molecule type: protein
A/Residues: 39-378 <EN>
C/Comment: This enzyme is a copper-containing enzyme. It is involved in the alternative C/reduces nitrite ion to nitric oxide, and it receives electrons from the blue copper C/genetics:

```
C;Keywords: copper; electron transfer; oxidoreductase
F;1-38/Domain: signal sequence #status predicted <SIG>
;39-378/Product: nitrite reductase #status predicted <MAT>
.
```

Query Match	17.64%; Score 461.5; DB 2; Length 378;
-------------	--

Best Local Similarity 36.3%; Pred. No. 5.8e-24;
Matches 134; Conservative 45; Mismatches 149; Indels 41; Gaps 12;

QY 19 MLCGCSNOADKKAQPKSSTVDAAKTANADNAAOEHOGELPVITDAIVTHAPVPPVDR 78
DB 12 MLAGALAG--AVAPLHTQAAHAGAAAAGAAPDISLPRVKDL-----VKPPFVH 64
QY 79 DH-----PAKVVKMETVEKVMRL-ADGVEYQFWTFGGQVPGQMIRVREGDTIEVQFS 130
DB 65 AHDQVAKTGPRVVEFTMTTEKKLVIREGTEIHAFTFNGSVGPIMVHENDYVELRLI 124
QY 131 NHDSKMPHNVDFHATPGGGAEMASTARHSTSFRLQGLVYVHCAPVCM--- 187
DB 125 NPDNTLLNHIDHAAATGALGGALTVQNGEETTLRFKATKPGVFVYHC--APEGVVPW 182
QY 188 HIANGVYGLIVEPKGLP-----KVDEYVVMQDFFYTK-----GKYGEQGLQFP 233
DB 183 HVTSSNGAAMVLPRLGLDEKQPLTYDKITVYVGEQDFVPPDEAGNTKKTETPEAYE 242
QY 234 DMEKAIRE-DAEYVNGSVGALTGENALAKAVGETVRLFVNGGPNLTSSFFHVEIFD 292
DB 243 DAVKAKRTLPHTHIVFNGAVGALTGDHALTPAAVGE--RVLVHVSQANRDRPHLIGSHG 300
QY 293 KYHFEKGKGNH---NIQTLLIPAGGAATFEKVDVPGDVLVDHAIFFAFKNGALGILK 349
DB 301 YV-WATGKFRNPDPDLOETWTLIPGTAGAAFYTFROPGVYAVVNHILIEAFELGAAGHFK 359
QY 350 VEGENHEI 358
DB 360 VTGEMNDL 368

RESULT 5
A48936
nitrite reductase (EC 1.7.99.3) - Pseudomonas sp.

C:Species: Pseudomonas sp.
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48936
R:Ye, R.W.; Fries, M.R.; Bezborodnikov, S.G.; Averill, B.A.; Tiedje, J.M.
Appl. Environ. Microbiol. 59, 250-254, 1993
A:Title: Characterization of the structural gene encoding a copper-containing nitrite reductase from Pseudomonas sp. strain A48936
A:Reference number: A48936; PMID:93175864; PMID:8439151
A:Accession: A48936
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-379 <YE1>
A:Cross-references: GB:M97294; NID:g151396; PIDN:AACT9132.1; PID:g151397
A:Experimental source: Isolate G-179
A:Note: sequence extracted from NCBI backbone (NCBIN:125554, NCBIIP:125555)
C:Keywords: copper; oxidoreductase

Query Match 17.3%; Score 453; DB 2; Length 379;

Best Local Similarity 37.3%; Pred. No. 2.2e-23;
Matches 118; Conservative 45; Mismatches 119; Indels 34; Gaps 10;

QY 72 VPPVVDNRD-----PAKVVKMETVEKVMRLAD-GVEYQFWTFGGQVPGQMIRVREGD 123
DB 59 VKRPFVHATQKAEGRPKVVEFKMTIOEKKIIVDDKTEVHNTFDGVSVPGRPMIYHDD 118
QY 124 TIEVQFSNHPDSKMPHNVDFHATPGGGAEMASTARHSTSFRLQGLVYVHCAPVCM--- 183
DB 119 YVELTLVNPDTMELQHNIIDFHSATGALGGALTVNPGDVAVRFKATKAGVFVYHC--A 176
QY 184 PYGM---HIANGVYGLIVEPKGLP-----KVDEYVVMQDFFYTK-----GKYG 226
DB 177 PEGMVPWHVTSNGMGAIMVLPRLGLDKHGLVYKVVYVGGQDFVPPDEAGNKFKKYE 236
QY 227 EQLQFPDEKAIRE-DAEYVNGSVGALTGENALAKAVGETVRLFVNGGPNLTSSFFH 285
DB 237 SAGEAIPDVLKANKITLPHIVFNGAVGALTGDHALTPAAVGE--RVLVHVSQANRDRPH 294
QY 286 VIGETIDKVFEGKGNH---NIQTLLIPAGGAATFEKVDVPGDVLVDHAIFFAFK 342

DB 295 LIGHGDIY-WATGKFRNPDPDLOETWTLIPGTAGAAFYTFROPGVYAVVNHILIEAFEL 353
QY 343 GALGILKVEGENHEI 358
DB 354 GAAGHFKVYGDWMDL 369

RESULT 6

QY0170
nitrite reductase (EC 1.7.99.3) - Alcaligenes denitrificans subsp. xylosoxydans
C:Species: Alcaligenes denitrificans subsp. xylosoxydans
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0170
R:Suzuki, E.; Horikoshi, N.; Kohzuma, T.
Biochem. Biophys. Res. Commun. 255, 427-431, 1999
A:Title: Cloning, sequencing, and transcriptional studies of the gene encoding copper-catalyzed nitrite reductase from Alcaligenes denitrificans subsp. xylosoxydans
A:Reference number: JG0170; PMID:99160880; PMID:10049725
A:Accession: JG0170
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <SUZ>
A:Cross-references: GB:AF051831
C:Keywords: copper; oxidoreductase

Query Match 17.2%; Score 451; DB 2; Length 360;

Best Local Similarity 35.8%; Pred. No. 2.8e-23;
Matches 134; Conservative 45; Mismatches 135; Indels 60; Gaps 15;

QY 41 AAKTANADNAAOEHOGELPVITDAIVTHAPVPP--PVDRHAPKVVKMETVEKVMRLA 98
DB 18 AAGTMAODA-----DKLPHTKVLVAPPOVPHQATKSPKVVEFTMTTEKKVID 71
QY 99 D-GVEYQFWTFGGQVPGQMIRVREGDTIEVQFSNHPDSKMPHNVDFHATPGGGAEMAF 157
DB 72 DKGITLQANTFNGSNPFPVTVVHEDVYDVLVNPATNMPHNVDFHATGALGAKLTN 131
QY 158 TAPGHTSTFRLQGLVYVHCAPVCM---HIANGVYGLIVEPKGLP-----K 207
DB 132 VNPGEQATLRFADSGTGVYHC--APEGMVPMHIVSGMSTGLMVLPRGLADPQGRPLH 189
QY 208 VKKEYVVMQDFFY---TKGKGEGL--QFPDEKAIRE-DAEYVNGSVGALTGEN 259
DB 190 YDRAVYTFEFDYIRKPGDKYATLAESYGDVQVWRVTLTPSHIVNGVGAALTGAN 249
QY 260 ALKAVGETVRLFVNGGPNLTSSFFHVEIFD-KYHFEKGKGNH---NIQTLLIPAG-- 314
DB 250 ALTKAVGETVLI--HSQANRDRPHLIGHGDMV-WETGKFRNPDPDLOETWTLIPRGSA 306
QY 315 GAATFEKVDVPGDVLVDHAIFFAFKNGALGILKVEGENHEIYSHKOTDAVYLPREGAP 374
DB 307 GAALYTFK--QPGVYAVVNHILIEAFELGAAGHIVGKMD----- 346
QY 375 QALDPOEAPKTPAP 388
DB 347 ---DLMKQIKAPAP 357

RESULT 7

S32112
(EC 1.7.2.1) (EC 1.7.2.1) - Pseudomonas aureofaciens
C:Species: Pseudomonas aureofaciens
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Jun-2002
C:Accession: S32112
R:Zumft, W.G.
submitted to the EMBL Data Library, March 1993
A:Reference number: S32112
A:Accession: S32112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-363 <ZUM>
A:Cross-references: EMBL:Z21945; NID:g287906; PIDN:CAA79939.1; PID:g287907
C:Keywords: copper; oxidoreductase

A:Residues: 1-455 <KUR>
A:Cross-references: GB:AE007870, PIDN:AAK89058.1, PID:g15158854, GSPDB:GN00170
C:Species: Sinorhizobium meliloti
A:Gene: AGR_L_970
A:Map position: linear chromosome

Query Match 16.8%; Score 442; DB 2; Length 455;
Best Local Similarity 33.2%; Pred. No. 1,6e-22;
Matches 122; Conservative 58; Mismatches 131; Indels 56; Gaps 12;

21 SGCNQADKAAKPSSTVDAAKTANAD-----NAAOEHOGELPVIDAIVTHAEPV 73
DB ACAAEKKAAPKPLTSAEIALPRAKVDLPKPPVHAHQKAGGPKVEFTLT----- 159
QY 74 PPVDRHAPKAVVKKMEVEKVRADL-GVEYQWTFEGGQVPGOMIRVREGDTIEVOFSNH 132
DB 160 -----IK-----EQKMLDDKGTVEHAMFNSVSPGPLMVHDDVVELTLINP 203
QY 133 PDSKAPHVDFPAATGPGGGAASFTAPGHTSTFSFKALQPLYYHCAVAPVGM--HI 189
DB 204 DTNELQHNIDFHSAGALGGGGGLTYVNPGEKAILRFKAKAGVFVYHC--APPGVMPMHV 261
QY 190 ANGWYGLIVEKEGLPK-----VDKEYVMQGFYTK-----GKYGEGLQPFDM 235
DB 262 TSGMGAIVLDPREGLTGDKGKELYVDKYVYVGBODFYIPRDENGPFKYESAGAMADT 321
QY 236 EKAIKE-DAEYVFNQSVGALTGEMALKAQVETVRLFPVNGGPNLTSSFVHGEIFDKV 294
DB 332 LEVWRKLPSSHIVFNGAVGALTGEMALQAAVE--KVLIVHSQANRDRPHLIGHGDDV 379
QY 295 HPEGKGENH--NIQTLLIPAGAAITFEKVDVPGDYLVDHAIFFRAFNKALGILKYE 351
DB 380 -WATGKFRNPRLDDETFIPGDTAGAAAYTFEGGIVAYVNHNLIEAFELGAAAHFKVT 438
QY 352 GEENHEI 358
DB 439 GEMNNTL 445

RESULT 11
A95347
probable nitrite reductase (EC 1.7.99.3) Nirk [imported] - Sinorhizobium meliloti (strain C) [Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001
C:Accession: A95347
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bove, .; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yen, K.C. Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti A:Reference number: A95347; MUID:21396509; PMID:11481432
A:Accession: A95347
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <KUR>
A:Cross-references: GB:AE006469, PIDN:AAK65339.1, PID:g14523797, GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Pihan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chait, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: nirk
A:Genome: plasmid
C:Keywords: oxidoreductase

Query Match 16.7%; Score 439.5; DB 2; Length 376;
Best Local Similarity 33.2%; Pred. No. 1.8e-22;
Matches 121; Conservative 52; Mismatches 128; Indels 63; Gaps 13;

QY 60 PVIDAIVTHAP-----VPPVDRDHP-----AKVVKKETVE- 92
DB 24 PLIGAVSAHAEBAVAKTAHINVASLPRVKVDLPKPPVHAHQKAGGPKVEFTLTIEE 83
QY 93 -KWRMLADGVEYQWTFEGGQVPGOMIRVREGDTIEVOFSNHDSKMPHNVDFHAATGPGG 151
DB 84 KKIIVDEQGTIELHANTFNSVPGPLMVHDDVVELTLINPTNTLQHNIDFHSATGALG 143
QY 152 GAESAFTAPGHTSTFSFKALQPLYYHCAVAPVGM--HIANGWYGLIVEPKGLP- 206
DB 144 GGLTIVNPBGDTIVLRFPAKAGVFVYHC--APPGVMPMHVTSGNMGALVLPREGLTG 201
QY 207 -----KVDKEYVMQGFY-----TKGY-----GEGLQPFDMKAIREDAEYVFNFS 251
DB 202 KGNSTIVKYVYVGBODFYVRDANGKPKKYESGEAADLEWRTL--TSHIVFNCA 259
QY 252 VGALTGENALKAQVETVRLFPVNGGPNLTSSFVHGEIFDKVHPEGKGENH--NIQT 308
DB 260 VGALTGDSALKAQAVE--KVLIVHSQANRDRPHLIGHGDDV-WATGFRNAPVDQET 316
QY 309 TLIPAGGAITFEKVDVPGDYLVDHAIFFRAFNKALGILKEGEENHEIYSHKQTDAY 368
DB 317 WFTPEGTAGAAAYTFEGGIVAYVNHNLIEAFELGAAAHFAVTDGMNDLMT-----SVR 371
QY 369 LPEG 372
DB 372 APSG 375

RESULT 12
AC3633
nitrite reductase (EC 1.7.99.3) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AC3633
R:DelVecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, .; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leless Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis A:Reference number: AD3252; PMID:11756688
A:Accession: AC3633
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <KUR>
A:Cross-references: GB:AE008918, PIDN:AAL54230.1, PID:g17985202, GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10988
A:Map position: II
C:Keywords: oxidoreductase

Query Match 16.4%; Score 431; DB 2; Length 376;
Best Local Similarity 34.8%; Pred. No. 6.9e-22;
Matches 122; Conservative 48; Mismatches 147; Indels 34; Gaps 11;

QY 43 KTAANDNAAOEHOGELPVIDAIVTHAEPVPPVDRHAPKAVVKKMEVEKVRADL-GV 101
DB 38 RKASLEEIALPRLQ-KVELVDPPIVHA--HTQVAEGGPKVQFTVIEKKIIVDDAGT 93
QY 102 EYQWTFEGGQVPGOMIRVREGDTIEVOFSNHDSKMPHNVDFHAATGCGGAASFTAPG 161
DB 94 EVHAMTFNGVTPGPLMVHDDVVELTLINPTNTLQHNIDFHAATGALGGGGLTEINPG 153
QY 162 HTSTFSFKALQPLYYHCAVAPVGM--HIANGWYGLIVEKEGLP-----KVDKE 211
DB 154 EKTIVLRFKATKPGVFVYHC--APPGVMPMHVTSGNMGAVVLPREGLTGDKGNKLTVDKV 211
QY 212 YVWQGFYTK-----GKYGEGLQPFDMKAIKE-DAEYVFNQSVGALTGEMALKA 263
DB 212 YVGSQDFYVRDENGKVTYAPGDAVEDVTKWRTLTPPHVFNQSVGALTGKALTA 271
QY 264 KVGETVRLFPVNGGPNLTSSFVHGEIFDKVHPEGKGENHNI--QTTLLIPAGGAITFE 321

Db 272 KXGKVLII--HSQANDRPHLIGHGQYVWATGKFNTPPVDDGTWTFPGGAHAARY 329
 QY 322 KVDVPGDYVLVDHAFRAFNKAGLILKVEGENHEIYSHKOTDAVYLPEG 372
 Db 330 TERPGIYAVVNHNLIEAFELGAAAHFYKTEWMDLMT-----SILAPSG 375

RESULT 13

139582
 nitrile reductase (EC 1.7.99.3) [validated] - Alcaligenes faecalis
 C:Species: Alcaligenes faecalis
 C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #ext_change 15-Sep-2000
 C:Accession: J39582
 R:Nibelkuma, M.; Suzuki, J.; Kukimoto, M.; Ohnuki, T.; Horiuchi, S.; Beppu, T.
 J. Gen. Microbiol. 139, 725-733, 1993
 A:Title: Cloning and characterization of nitrile reductase gene from Alcaligenes faecalis
 A:Reference number: J39582; PMID:93294530; PMID:8515232
 A:Accession: J39582
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-376 <RES>
 A:Cross-references: GB:D13155; NID:G398121; PIDN:BA02440.1; PID:G398122
 A:Gene: nitr
 C:Keywords: oxidoreductase
 F:311,172,181,186/Binding site: copper (His, Cys, His, Met) (type 1) #status experimental
 F:316,171/Binding site: copper (His) (type 2) (shared with trimeric partner 2) #status experimental
 F:342/Binding site: copper (His) (type 2) (shared with trimeric partner 1) #status experimental

Query Match 16.3%; Score 427.5; DB 2; Length 376;
 Best Local Similarity 33.7%; Pred. No. 1.2e-21;

Matches 129; Conservative 51; Mismatches 158; Indels 45; Gaps 14;

QY 15 LSLMLSGCSNQADKAAQPKSSTVDA---AAKTANADNAASQEHGELPVIDAIVTHAP 70
 Db 13 LAGALAG-----ALAPVLATTSAMGQAVRKATTAETALPRQ-KVELVDPFVHAH 64
 QY 71 EVPPVDRDHPAAVYVVKMEVEKVMRLAD-GVEYQFTFGQVPGCMIRREDTIEVQE 129
 Db 65 S---QVAEGGPKVVEFTVVEEKIYVDDAGTEVHMAAFNGVTPGPMVHHODVLELT 121
 QY 130 SNHPDSMPHNVDFHAATGPGGGAASFTAPGHTSTFSFKALQGLLYVYHCAVAPGM-- 187
 Db 122 INEPTTILMNIIFHAATGALGGGGLTEINPEKTIIRFQATRGVYVHC--APPGMV 179
 QY 188 -HIANGVGLIVEPEKLP-----KVDKEYVVMGDFYT---KGKYE--QGLQ 232
 Db 180 MHWVSGNKAIVLPRGLHDGKGAALTYKIVYVGEQDFVPRDEGKYYEAPQDAY 239
 QY 233 FDMEKALRE-DAEYVNVSGVAGLTGENALKAVGETVRLFVNGGPNLTSSFRVIGEIF 291
 Db 240 EDTVAKVARTLPTHTVFNAGVAGALTGDKAMTAVGE--KVLIVHQAQNRTPRHLLIGG 297
 QY 292 DKVHPEGKGENNI--QTLIPAGCAITEFKVDVGVLDVLDHAFRAFNKAGLILK 349
 Db 298 DYVWATGKFNTPPVDDGTWTFPGGAHAARYTQDPGIYAVVNHNLIEAFELGAAAHK 357
 QY 350 VEGEENHEIYSHKOTDAVYLPEG 372
 Db 358 VTGEMNDLMT-----SVLAPSG 375

RESULT 14

E84274
 membrane protein (imported) - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84274
 R:Ng, W.V.; Kennedy, S.P.; Mahirae, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitner, B.; Keller, K.; Cruz, R.; Danon, M.J.; Hough, D.W.; Maddocks, D.G.; Jabic
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; PMID:20504483; PMID:11016950

A:Accession: E84274
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-379 <STO>
 A:Cross-references: GB:AE004437; NID:G10580721; PIDN:AA019561.1; GSPDB:GN00138
 C:Gene: pan1

Query Match

Best Local Similarity 25.3%; Score 250.5; DB 2; Length 379;
 Matches 94; Conservative 52; Mismatches 133; Indels 93; Gaps 20;

QY 20 LSGCSNQADKAAQPKSSTVDAAK---TAADNADNAASQEHGELPVIDAIVTHAPEVPP 75
 Db 29 IAGCT--APDGEVSTDTTAAKOSGLPTSPPEVDATEGQNVTLKSV-----PA 78
 QY 76 VDRDHPAAVY--VMEETVEKVMRLADGVEYQFTFGGO-----VPGCMIRVREGDTIEV 127
 Db 79 VADVHPDLSMGQPVKLPV-----WAFATEDGDPSPGPIVREBQDIEV 124
 QY 128 QPSNHPDSMPHNVDFHAATGPGGGAASF-----TADGHTSTFSFKALQGL 175
 Db 125 TLDN-TDGRPPTLHFH-----GSQTAMEDDGVPTTGLRVGGEKHTYIIPANVGT 176
 QY 176 VYHCAVAPVGMHIANGVGLIVEPEKGLPRVDEYVVMQDFTTKKYGSGQIQPDM 235
 Db 177 HLYHCHY-OTQRHIDMGVGIIFRIDPK-GYEPADKXETFTVVDWDSLRNMA-- 228
 QY 236 EKAIREDAPVY--VNGSVGALTGENA-----LKAIVGETVRLFVNGGPNLTS 282
 Db 229 -----EDVDYSRTNPNPVFTYNGKSAPRTLHPEDGSPVIEGQTVLHLVNGG-YNMH 282
 QY 283 SFHVGELFDKVFEGGK--GENNIO-TLLIPAGCAITEFKVDV-PGDIYL---VD 333
 Db 283 PLHINHRFRVYKXGQGPAAQHTMTDTPMAPEHRTI-EFOADAPGIVLHMHKVN 341
 QY 334 HAFRAFNKAGL 345
 Db 342 HVMNGTFYPGGM 353

RESULT 15

AG2298
 hypothetical protein a113942 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AG2298

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriuch

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; PMID:21595285; PMID:11759840

A:Accession: AG2298

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <NR>

A:Cross-references: GB:BA000019; PIDN:BA075644.1; PID:G17133076; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Gene: a113942

Query Match

Best Local Similarity 9.5%; Score 248.5; DB 2; Length 338;
 Matches 76; Conservative 42; Mismatches 116; Indels 39; Gaps 12;

QY 94 VNRLAGVEYQFTFGQVPGCMIRREDTIEVQSNHPDSMPHNVDFHAATGSGGA 153
 Db 93 VIOLNAGVSNYINWDLNIGRIPIGLRAKQDRIIRVLF--HNQAGHSHSLHFHGV---HPA 146

QY 154 EASFAP---GHTSFESFKALOPGLVYHCANAPYGMIIANGMGLIIVEKEGJPKYD 210
 Db 147 EMDGRPRSPNNSSATIYEEFDAERYGHLHYCHIEPTTRIIAGLGYMFIIDEPTTPRPAD 206
 QY 211 EYVWQGDQFYTKGKYGEQLOPFDMEKARIDBAEYVWNGSVGLTGENALKAKGVETR 270
 Db 207 IVLWAG-----YDVDSNSHD--FYANGLPRHHY-DNP:QIQONQIR 248
 QY 271 LFVGN---GGRNLTSFHVIGEIFPKVHPEGGKGENNHIQTTLLPAGAA--ITEFKYDV 325
 Db 249 LYNVNIIEYDAVT--FHIHANFPDVRY--GMSKMASEKTDVITMGVAERHIIIEFARY 304
 QY 326 PGDYVLVDHALFRAFNKGALGILKY--EGEENH 356
 Db 305 PGKTFHFPHODALAEH--GCMGQFEVYVANNQNH 336

RESULT 16

B81167
cytochrome c552 NMB0717 precursor [similarity] - Neisseria meningitidis (strain MC58 serotype C)
C|Species: Neisseria meningitidis
C|Date: 01-Sep-2000 #sequence_revision 01-Sep-2000 #text_change 19-Jan-2001
C|Accession: B81167
R|Rettlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
rt, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Mastignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A|Authors: Grandi, G.; Sun, L.; Smith, H.O.; Frazer, C.M.; Moxon, E.R.; Rappunli, R.; Varki,
A.|Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A|Reference number: AB1000, MUID:20175755; PMID:10710307

A:Molecule type: DNA
A:Residues: 1-138 <TER>
A:Cross-references: GB:AE002426; GB:AE002098; NID:g7225939; PIDN:AAF41130.1; PID:g722594
A:Experimental source: serogroup B, strain MCS8
C:Genetics:
A:Gene: NMB0717
C:Superfamily: T. aquaticus cytochrome c552; cytochrome c6 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-138/Product: cytochrome c552 #status predicted <MAT>
F:33-114/Domain: cytochrome c6 homology <CY6>
F:43_46/Binding site: heme (Cys) (covalent) #status predicted
F:47_97/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match	8.5%	Score 222;	DB 1;	Length 138;
Best Local Similarity	42.7%	Pred. No.3.1e-08;		
Matches	44;	Conservative 14;	Mismatches 45;	Indels 0;
			Gaps 0;	
Oy	398	GCATDSCACACHQDDGKGVNAPPLANSPLYLNADHRAASIVANGLSGKITNGNGYE	457	
		: : :		
Db	35	GQKRYESCVACHGKKGEGRGTMPEPLYSDFIMKKPOVLLHSWVGKINGITKNGKTYN	94	
Oy	458	SVMPAIALSDQOIANVITYTNSFCNKGGSADDAVAKAKTK	500	
		: : :		
Db	95	GFMPETALSDADIAAVATYIMNAEDNGSGSVTEKDVQAKSKK	137	

RESULT 17
F81938
cytochrome c552 NMA0925 precursor [similarity] - Neisseria meningitidis (strain Z2491)
C:Species: Neisseria meningitidis
C:Date: 01-Sep-2000 #sequence_revision 01-Sep-2000 #text_change 02-Feb-2001
C:Accession: F81938
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
R:Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajadaram
Nature 404, 502-506, 2000.
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: F81938
A:Molecule type: DNA
A:Residues: 1-163 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84197.1; PID:g737963

```

A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: MMA0919; MMA0925
C:Superfamily: T. aquaticus cytochrome c552; cytochrome c6 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein
F.1-21/Domain: signal sequence #status predicted <SIG>
F.122-163/Product: cytochrome c552 #status predicted <MAT>
F.128-109/Domain: cytochrome c6 homology <CY6>
F.318-41/Binding site: heme (Cys) (covalent) #status predicted
F.142-92/Binding site: heme (Iron) (His, Met) (axial ligands) #status predicted

Query Match      8.5%; Score 222; DB 1; Length 163;
Best Local Similarity 42.7%; Pred. No. 3.9e-08;
Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0

QY      398 GKATYDSNCAQHOPDGKGVNAPPLANSYLYNADHRAASIVANGSLGKITVNGOYE 457
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       30 GQKYVESNCVACHKKGGEGRGTEFPPLRYRSPFMKKKQVLLHSWVGINGITIKVNGKTYN 89
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      458 SVMPAIALSDQIANVITYTNSGNGKGGASADPAVAKAKTK 500
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       90 GFMPAIALSDADIAAVATYINNAADNGGGSVTEKDVKQAKKKK 132

```

RESULT 18
552253
copper resistance protein proA precursor - Escherichia coli plasmid pRJ1004
N:Alternate names: copper-binding protein
C:Species: Escherichia coli
C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 11-Jun-1999
C:Accession: S70159, S52253
R:Brown, N.L.; Barrett, S.R.; Camakaris, J.; Lee, B.T.O.; Rouch, D.A.

A:Title: Molecular genetics and transport analysis of the copper-resistance determinant
A:Reference number: S70159; MUID:96130847; PMID:8594334
A:Accession: S70159
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-605 <BRO>
A:Cross-references: EMBL:X83541; NID:G619126; PIDN:CMA58525.1; PID:G619128
C:Genetics:
A:Gene: pcoa
A:Genome: plasmid pRJ1004
C:Superfamily: laccase
C:Keywords: copper binding
F1-32/Domains: signal sequence #status predicted <SIG>
F13-605/Product: copper resistance protein pcoa #status predicted <MAT>

[illegible]

Db 337 RTGVARTLATRGLSLAAV-----PRDPRRLTWMEDGKMGMDHDAAGMDHSGMGMDNS 392
 QY 407 AACHPDCKGVN---AFPLPANSDYLNADHAPAAST 440
 Db 393 GEMMSMDGADLPDSGTSSAPMDHSHSMAGMDHSHWAGM 429

RESULT 19
 B83910
 hypothetical protein B83910 [imported] - Bacillus halodurans (strain C-125)
 C/Species: Bacillus halodurans
 C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C/Accession: B83910
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hara
 Nucleic Acids Res. 28, 4317-4331, 2000
 A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A/Reference number: A83650; MUID:20512582; PMID:11058132
 A/Accession: B83910
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-500 <STO>
 A/Cross-references: GB:AP001514; GB:BA000004; NID:910174613; PIDN:BA05801.1; GSPDB:GN00
 A/Experimental source: strain C-125
 C/Genetics:
 A/Genes: B82082

Query Match 7.4%; Score 194; DB 2; Length 500;
 Best Local Similarity 20.7%; Pred. No. 1.4e-05;
 Matches 105; Conservative 71; Mismatches 198; Indels 134; Gaps 23;

QY 20 LSGCSNQADKAAPKSTVDAAKTAADNAASOEHGELPVIDAIVTHAPEVPPDRD 79
 Db 18 IAGCS-----DHSMDHGNQODEQNEGT----- 44

QY 80 HPAKVVKMETVEKVKRLADGVEYQFTFGGOVPGQWIRVEDDTLEVQSNHPSKMP 139
 Db 45 EDGKVVNLEVTENHMFNEVMDATVNGTLPQGEIRVOEGDEVIVANK--SLNVPT 102

QY 140 NVDPHAATPG---GGAESTAPAGHTSTFSKALQPLYYHCALVAVGNIHANGY 194
 Db 103 ALHIGFPPVNEMDGVVQNAIMPEEFTVEYQANVPQTYWH--SHDQATQVNRGLY 161

QY 195 GLTIVEKEGLPKYDKXY-----VMGDPYTKG-----KXGE---OGIQRPMDE 236
 Db 162 GVIVET-----VDPESTDTDAIVIDEFASMGCMGMDHSHMDHMDHMDHSHS 215

QY 237 KAIREDAEYVFNFGVGLTGENALKAKVETVRLFVNGGPNLTSSFHVIGEIFDKVNF 296
 Db 216 EMNNEMVDTVINGKADPQI--ESINNVESGRVRLRVNAG--LFTQVVSIPESHFKLTHY 272

QY 297 EGGKGENHNIQTLLIPAGAAITEFKVDVPEDY---VLVDHAFRANKGALGLKKEGE 353
 Db 273 DGGPVPNEPVDLQD-----TAFRIAPARYDLLEIIMD-----NPGAWGI--QVFAE 315

QY 354 ENNE-----YSHKQDAVVLPEAPALIDQAPKTPAPANLQEQKAGKAT--YD 403
 Db 316 ENQRRLOALIPLYEYEDDEL-----QTVDSISFFDLTYGILKELDIDGITKEYD 368

QY 404 SNCAACHQPD---GKGVNP-----AFPLPANSDY-----LNADHAPAAST 441
 Db 369 MLGSDGCGDTFTINGKKFPHHELYDVEBGMVFTIINDTDPHMHMLGDPFHVIVSKG 428

QY 442 ANGLSG---KITVN---GNQYESVMPA 462
 Db 429 GKAIQGSPLVKDTLNVPRGETYEIVFEA 456

RESULT 20
 A98239
 hypothetical protein AGR_L_1722 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C/Species: Agrobacterium tumefaciens
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002

C/Accession: A98239
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Mollam, C.; Allington, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.,
 Science 294, 2323-2328, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A/Reference number: A97359; MUID:21608551; PMID:11743194
 A/Accession: A98239
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-449 <KUR>
 A/Cross-references: GB:AE007870; PIDN:AAK89435.1; PID:915159296; GSPDB:GN00170
 C/Genetics:
 A/Genes: AGR_L_1722
 A/Map position: linear chromosome

Query Match 7.1%; Score 187; DB 2; Length 449;
 Best Local Similarity 21.2%; Pred. No. 3.6e-05;
 Matches 97; Conservative 63; Mismatches 175; Indels 122; Gaps 20;

QY 12 ICALSLMLSGCSNQADKAAPKSTVDAAKTAADNAASOEHGELPVIDAIVTHAPE 71
 Db 7 LIGVSLAAWSSA-----AMATSSNS-----LPEALMETATQAP- 42

QY 72 VPPVDRDHPAKVVK-----METVEKVKRLADGVEYQFTFGGQVP 113
 Db 43 VKPTGPDYNPVVTINGWTLPPRMNNGVKEFHLVAEPVERM--AEGMTARLMGYGQSP 100

QY 114 GQMIRVREDDTLEVQSNHPSKMPHN--VDHFAATPG---GGAESTAPAGHTSNF 166
 Db 101 GPTIEAVEEDRVRIFVTN---KLPEHTTIHNGMILPBGMDGVGLSOPHI PVGKTFVY 156

QY 167 SFKALQPLYYHCALVAVGNIHANGYGLTIVEPE--GLPKYDKXYVMQDFFYTKGY 225
 Db 157 EFDLVKSGTFMNH--PHSDMVGMAQMGKFFIHPDAPFMVVDHDFVLNAF----- 209

QY 226 GEGGLQPFMEKAIREDAYVFNFGVGLTGENALKAKVETVRLFVNGGPNLTSSPH 285
 Db 210 ---DINPGASVPKIMTMTDFNLMAWNSRVFPGIDPLVSKNDRVVRVG---NLMTNH 262

QY 286 VI---GEIDKXNHEG--KGENHNIQTLL--IPAGAAITEFKVDVPEDYVLVDHAFRA 339
 Db 263 PLHMGYDEVTCTDGVWRPAPRPEVSIDIPVGMARAYEDAKYVGMALHCHSHHT 322

QY 340 FNKGALGILKVEEENHEIYSHKQDAV-----YLPEGAPOAID---TOEAP 383
 Db 323 KN--AMG-----HDVPTFGVDKKTVEKIRKIRDPYMPGTAGMADGAMEMELP 371

QY 384 KTRAP-----ANLQEQIAGKAT 401
 Db 372 ENTVPMTGMPGHPHPIEMGMEFVVKREGISAGDYT 408

RESULT 21
 AC3047
 multicopper oxidase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C/Species: Agrobacterium tumefaciens
 C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C/Accession: AC3047
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
 erage, G.; Gille, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AC3047
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-449 <KUR>
 A/Cross-references: GB:AE008669; PIDN:AA44793.1; PID:917742433; GSPDB:GN00187
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:

A:Gene: COPA
A:Map position: linear chromosome

Query Match 7.1%; Score 187; DB 2; Length 449;
Best Local Similarity 21.2%; Pred. No. 3.6e-05;
Matches 97; Conservative 63; Mismatches 175; Indels 122; Gaps 20;

QY 12 ICALSNALMISGSGSNADKAAQPKSSTVDAATAKANDNNAOEHOGELPVIDAIVTHAPE 71
DB 7 LIGSAAMVSSA-----AMAKTSNS-----LPEALMETATQAQ-- 42
QY 72 VEPVPDRDHPAKVVK-----METVKWRLADGVYQFTFGGQVP 113
DB 43 VEPPTGPDYNPVVTNGMTLPFRMNNVKEFHLVAPVRREM--AEGMTARLMGVNGOSP 100
QY 114 GQMIRVREGDTIEVQSNHPSDKMPHN--VDFHAATGPG-----GGAASFTAPGHTSPF 166
DB 101 GPTIEAVEGDRIKRIEFTN--KLPEHTTIHMGMLPBGMDGVGLSQPHIPVGTFTVY 156
QY 167 SFKALQPLLYVHCAPVGMHANGMYGLIIVEPKR--GLPKVKEYVMQDPFYTKGKY 225
DB 157 EFDLVKSGTFMYH--PHSDEMVOAMGMGFVYIHPKDPAFMPYDRDFVFLNAF----- 209
QY 226 GEGGQIPFMKAIREDAYVVFNGSVGALTGENALKAQVETVRLFVNGGPNLTSSPH 285
DB 210 --DINPGASVPKIMTMDPFLMAMNSRVFPGIDPLVSKNDRVVRVG---NLMTNH 262
QY 286 VI---GEIPDKVHEGEG--KGENHNIOQLL--IPAGSALTEPFKVDVPGDYVLVDHAI 339
DB 263 FTHMGVDEVTCTDGGWVRPEARPEVSIIDIPGAMRAYEDDAKYVGDMALHCHSKSHHT 322
QY 340 FNKGALGILKVEGENHEIYSHKQTDV-----YLPEGAPOAID---TOEAP 383
DB 323 NW--AMG-----HDVPTFGVDKKTVEKIRIRPDYMPMGAGNADGAMEMELP 371
QY 384 KTRPAP-----ANLQOIRAKKAT 401
DB 372 ENTVPMTGMPHGPHEMGMFSSVYVRREGISAGDVT 408

RESULT 22

C95332
Probable copper-containing oxidase [imported] - Sinorhizobium meliloti (strain 1021) mag
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: C95332
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Bariloy-Hubler, F.; Bows
Proc. Natl. Acad. Sci. U.S.A. 98, 9683-9688, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: C95332
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-449 <R>
A:Cross-references: GBR:AB006469; PIDN:AAK65221.1; PID:914523669; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Bariloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Author: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kles, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenberg, M.; Vorholter, F.U.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Gene: Sma1038
A:Gene: plasmid

Query Match 6.6%; Score 174.5; DB 2; Length 449;
Best Local Similarity 23.9%; Pred. No. 0.00025;
Matches 92; Conservative 55; Mismatches 175; Indels 63; Gaps 19;

QY 40 AAKTAN---ADNAAQO--EHOGELPVIDAIVTHAPEVPPVDRD-----HPAKVVK-- 87
DB 19 AAKTSMNGLPAAVMEATGQPLK-----PTSGPDYNPVVTNGMTLPFRMNNVKEF 73
QY 88 ---METVKWRLADGVYQFTFGGQVPGQIRVREGDTIEVQSNHPSDKMPHN--VD 142
DB 74 HLVAEPVRREM--ADDMRAYLMGVNGOSPGRPIEAVEGDRIKRIEFTN--KLPEHTTIH 127
QY 143 PHAATGPG-----GGAASFTAPGHTSPFALQPLGILYVHCAPVGMHANGMYGLI 197
DB 128 WHGMLPBGMDGVGLSQPHIPVGTFTVYEFVLVSGTFMYH--PHSDEMVOAMGMGF 186
QY 198 IVEPEKGLPK--VDKEYVMQDPFYTKGKEGQLOPFDEKAIREDAYVVFNGSVGA 254
DB 187 VHPHD--PKPEPVDRDVFLLNAF-----DIDGSVYPRIMETDFFNMCMNSRI 235
QY 255 LTGENALKAQVETVRLFVNGGPNLTSSPHVY--GEIPDKVHEGEG--KGENHNIOQT 309
DB 236 FPDIPPLVSKNDRVVRVG---NLMTNHPIHMGVDFEVTCTDGGWVRPEARPEV 291
QY 310 L--IPAGSALTEPFKVDVPGDYVLVDHAI FRAFNKGALGILKVEGENHEIYSH--KQTDV 367
DB 222 IDIPGAMRAYEDDAKYVGDMALHCHSKSHHTNMGHDIPFTFGVSKSVAKIKLRE 351
QY 368 YLPEGAPOAID---TOEAPKTRPAP 388
DB 352 YMPGKTGMADMGEMELPENTIP 376

RESULT 23

A51027
L-ascorbate oxidase (EC 1.10.3.3) [validated] - zucchini
N:Alternate names: ascorbase
C:Species: Cucurbita pepo var. melopepo (zucchini)
C:Date: 08-Nov-1989 #sequence_revision 14-Nov-1997 #text_change 15-Sep-2000
C:Accession: A51027; A30066
R:Messerschmidt, A.; Ladenstein, R.; Huber, R.
submitted to the Brookhaven Protein Data Bank, January 1992
A:Reference number: A51027; PDB:1A0Z
A:Accession: A51027
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA; protein
A:Residues: 1-552 <R>
R:Rossi, A.; Petruzzelli, R.
unpublished results, cited by Messerschmidt, A., Rossi, A., Ladenstein, R., Huber, R., B
9
A:Reference number: A30066
A:Contents: sequence; X-ray crystallography, 2.5 angstroms
A:Accession: A30066
A:Molecule type: protein
A:Residues: 52-67;100-108, 'H';445-454;498-517 <ROS>
R:Messerschmidt, A.; Ladenstein, R.; Huber, R.; Bolognesi, M.; Avigliano, L.; Petruzzelli,
J. Mol. Biol. 224, 179-205, 1992
A:Title: Refined crystal structure of ascorbate oxidase at 1.9 Angstroms resolution.
A:Reference number: A58657; MUID:92194315; PMID:1548698
A:Contents: annotation; X-ray crystallography, 1.9 angstroms
A>Note: the sequence reported in A51027 is attributed to R. Petruzzelli and A. Rossi
R:Messerschmidt, A.; Luecke, H.; Huber, R.
submitted to the Brookhaven Protein Data Bank, November 1992
A:Reference number: A51619; PDB:1A5O
A:Contents: annotation; X-ray crystallography, 2.2 angstroms, reduced form, residues 1-5
R:Messerschmidt, A.; Luecke, H.; Huber, R.
submitted to the Brookhaven Protein Data Bank, November 1992
A:Reference number: A51620; PDB:1A5P
A:Contents: annotation; X-ray crystallography, 2.59 angstroms, peroxide form, residues 1
R:Messerschmidt, A.; Luecke, H.; Huber, R.
submitted to the Brookhaven Protein Data Bank, November 1992
A:Reference number: A51621; PDB:1A5O
A:Contents: annotation; X-ray crystallography, 2.32 angstroms, azide form, residues 1-55
R:Messerschmidt, A.; Rossi, A.; Ladenstein, R.; Huber, R.; Bolognesi, M.; Gatti, G.; Mar
J. Mol. Biol. 206, 513-529, 1989
A:Title: X-ray crystal structure of the blue oxidase ascorbate oxidase from zucchini. An

R.Nieman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laid, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Kojima, J.; Kozlov, M.; White, O.; Salzberg, S.L.; Shapito, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A.Title: Complete Genome Sequence of *Calobacter crescentus*.
A.Reference number: A87249; MUID:21173698; PMID:11255647
A.Accession: H87368
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-570 <STO>
A.Cross-references: GB:AE005673; NID:g13422242; PIDN:AAK22948.1; GSPDB:GN00148
C.Genetics: CC0964

Query Match 6.2%; Score 163; DB 2; Length 570;
Best Local Similarity 25.2%; Pred. No. 0.0021;
Matches 69; Conservative 45; Mismatches 112; Indels 48; Gaps 14;

QY 73 PPVDRDHPAKVAVVMEYEVKVRMLADGVYQFWTFGGVPGQMIRVREGDTIEVQFSNH 132
DB 45 PPALSGEE-----IKL-TVGHMAKIDGKAGHAVVNGAIPQLIRLKEGQVRLSVTNT 98

QY 133 PDSKPHNVDFHAAACP---GGGASFTF--FGHTSTFSFKALQPGLYVYHCAVAPVGM 187
DB 99 LDEDT--SIHMGGLVPPQMDVPGVSPGIRGETFTYEPIRQSGTYWMH--SHSGL 153

QY 188 HIANGMYGLIIVEPK-EGLPKDYKEYVYVQCGDFTYTKGKGQGLQPFMEKAIREDAEVY 246
DB 154 QEQMHHYGMIIIDPAGEDVNAVDRHHVVLSDW-----SFLPHFELRLTKXSAGH- 204

QY 247 VFNGS----VGAITGEN-ALKAKVGETVRLFYNGGPNLTSSFFHVIIGEIFDKVHREGKG 301
DB 205 -FNROKQTVAGILKQDQSLKQVRA-----WGQRMPTDIDAVTVGSVYVYIANGHGPG 257

QY 302 EN-----NHIQTLIPAGGAATTEKVDVPG 327
DB 258 DNWTGLFAPGERVRLFTINAG--AMTTFVNRIPG 289

RESULT 27
S11027
L-ascorbate oxidase (EC 1.10.3.3) precursor - Cucurbita cv. Ebisu Nankin
C.Species: Cucurbita cv. Ebisu Nankin
C.Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 23-Sep-2002
C.Accession: S11027; S36936
R.Esaka, M.; Hattori, T.; Fujisawa, K.; Sakajo, S.; Asahi, T.
Eur. J. Biochem. 191, 537-541, 1990
A.Title: Molecular cloning and nucleotide sequence of full-length cDNA for ascorbate oxidase
A.Reference number: S11027; MUID:9031033; PMID:2143984
A.Accession: S11027
A.Molecule type: mRNA
A.Residues: 1-579 <EUR>
A.Cross-references: EMBL:X55779; NID:g18251; PIDN:CAA39300.1; PID:g18252
A.Accession: S36936
A.Molecule type: protein
A.Residues: 31-48 <ESA>
C.Superfamily: laccase
C.Keywords: oxidoreductase
F.373-568/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F.49-231/Distal site: copper (His) #status predicted
F.90-478/Binding site: copper (His) (type 2) #status predicted
F.192-134,480,536,538/Binding site: 2Cu-O cluster (His) (copper type 3) #status predicted
F.111-568/Distal site: #status predicted
F.193,397,473,542/Binding site: substrate (Trp, Trp, Glu, His) #status predicted
F.210-222/Distal site: #status predicted

Query Match 6.2%; Score 163; DB 2; Length 579;
Best Local Similarity 24.6%; Pred. No. 0.0021;
Matches 83; Conservative 44; Mismatches 130; Indels 80; Gaps 19;

QY 101 VEYQVWT-----FGGQVPGMIRVREGDTIEVQFSN--HPDSKPHNVDFHAAATG 148
DB 40 VEYMWADPCNENIVGINGQFPPTIRANAGDTIVVELINLHTEGVVH--WHGILQ 96

QY 149 PG-----GGAASFTF--FGHTSTFSFKALQPGLYVYHCAVAPVGMHIANGMYGLIIVEP 201
DB 97 RGTWADGTASISQCAINPGEFFNFVNDNGTFYH--GHLMQASAGLYGLIDP 153

QY 202 KEGLPK--VDKEYVMOGDFTYTKGKEQ-GLQPFMEKAIREDAEVYVNG- 250
DB 154 PQGKKEPPEHYDEINLLISDWCHQS IHKQEVLS--SKPIRWIGEPTILLNGRQFD 209

QY 251 -SVGALTGEN--ALKAKVGETVRLFYVNGGPNLTSSFFHVIIGEIFDKVHREGKGNNHIQ 307
DB 210 CSIAKYNSNLEPCKLKSGSEPCAPYI-----FVMPKTYRIRI-----AS 250

QY 308 TTLIPAGGAATTE---FKVDVPGDYV---LVDAIFPAFNKALGILKEGENHEIY- 359
DB 251 TTLAALNAPFALGNHRLVLEADGVYQFPTYSIDISGESVSLITTDQNPSENVYWSV 310

QY 360 ---SHKQTD---AVLPFGAQAIDTQEAAPTTPA 387
DB 311 GTRGRHPNTPPGLTLNLYLPNSVSK-LPTSPPEPTPA 346

RESULT 28
S66353
L-ascorbate oxidase (EC 1.10.3.3) precursor - common tobacco
C.Species: Nicotiana tabacum (common tobacco)
C.Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C.Accession: S66353
R.Kato, N.; Esaka, M.
Plant Mol. Biol. 30, 833-837, 1996
A.Title: cDNA cloning and gene expression of ascorbate oxidase in tobacco.
A.Reference number: S66353; MUID:96194464; PMID:8624413
A.Accession: S66353
A.Status: preliminary; nucleic acid sequence not shown
A.Molecule type: mRNA
A.Residues: 1-578 <KAT>
A.Cross-references: EMBL:D43624; NID:g599593; PIDN:BA07734.1; PID:g599594
C.Superfamily: laccase
C.Keywords: oxidoreductase
F.373-565/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F.87-475/Binding site: copper (His) (type 2) #status predicted
F.89,131,133,477,533,535/Binding site: 2Cu-O cluster (His) (copper type 3) #status predicted
F.108-565/Distal site: #status predicted
F.190,390,470,539/Binding site: substrate (Trp, Trp, Glu, His) #status predicted
F.207-221/Distal site: #status predicted

Query Match 6.2%; Score 162; DB 2; Length 578;
Best Local Similarity 22.9%; Pred. No. 0.0025;
Matches 104; Conservative 55; Mismatches 162; Indels 134; Gaps 23;

QY 91 VEKWRMLADGVYQFWTFGGVPGQMIRVREGDTIEVQFSN--HPDSKPHNVDFHAAATG 148
DB 37 VEYHMSPDGSESVVMGINGQFPPTIRAKGDTVAHVLNKLHTEGVVH--WHGIRQ 93

QY 149 PG-----GGAASFTF--FGHTSTFSFKALQPGLYVYHCAVAPVGMHIANGMYGLIIVEP 201
DB 94 IGTWADGTASISQCAINPGEFFNFVNDNGTFYH--GHLMQASAGLYGLIIVEP 150

QY 202 KEGLPK--VDKEYVMOGDFTYTKGKEQ-GLQPFMEKAIREDAEVYVNGVGLTGE 258
DB 151 GEGEKEPPEHYDEINLLISDWCHQS IHKQEVLS--SKPIRWIGEPTILLNGRQFD 209

QY 259 NALKAKVGETVRLFYVNGGPNLTSSFFHVIIGEIFDKVHREGKGNNHI- 306
DB 186 NPLR-WIDEPTILLNGRQYNCISLAFSPKPLPQCKLRGBOYAPQILRVPNKIYRL 244

QY 307 ---OTTLPAGGAATTEK---VDVPGDYVLDVAIFPAFNKALGILKEGENHEIYS 360
DB 245 RVASTLTLGSLIASLIGKHWVLEADGVY-----QPFs-----VQDMDIYS 286

QY 361 HKQTDVAVLPFGAQA---AIDTQ-EAPKTPAPANIQEQIKAGKATYDSCAACHQPDG 414
DB 287 GESYSVLEKTDQDPTKQWISINVRGEPKTPQGLTL-----LNYLPNSAS----- 332

QY 415 KGVNAPFPLAN--SDYLNADHARASIVANGLSGK-----ITVNG----- 453
 Db 333 -KFTLPPPLAPLWMDY---NHSKSPSNKIIPALMGSPKPPPPNRRIRITLLNTONKIDGYT 388
 QY 454 ---NQESVMPALALSDQOIANVITYTLNSFGNK 484
 Db 389 KMAINNVSLVLPF-----QVYGSINXGINAPFTK 418

RESULT 29

CCTW5T

cytochrome c552 [validated] - Thermus aquaticus

C:Species: Thermus aquaticus

C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 17-Nov-2000

C/Accession: A00112

R:Titani, K.; Ericsson, L.H.; Hon-nami, K.; Miyazawa, T.

B:Biochem. Biophys. Res. Commun. 128, 781-787, 1985

A:Title: Amino acid sequence of cytochrome c-552 from Thermus thermophilus HB8.

A:Reference number: A00112; PMID:85199131; PMID:2986626

A:Accession: A00112

A:Molecule type: protein

A:Residues: 1-131 <TIT>

A>Note: the source was designated as Thermus thermophilus

R:Than, M.E.; Hof, P.; Huber, R.; Bourenkov, G.P.; Bartunik, H.D.; Buse, G.; Soulimane,

submitted to the Protein Data Bank, June 1997

A:Reference number: A77356; PDB:1C52

A:Contents: annotation; X-ray crystallography, 1.28 angstroms, residues 1-131

R:Than, M.E.; Hof, P.; Huber, R.; Bourenkov, G.P.; Bartunik, H.D.; Buse, G.; Soulimane,

J. Mol. Biol. 271, 629-644, 1997

A:Title: Thermus thermophilus cytochrome c-552: a new highly thermostable cytochrome-c sd

A:Reference number: A59160; PMID:97428333; PMID:9281430

A:Contents: annotation; X-ray crystallography, 1.28 angstroms

C/Comment: This cytochrome appears to function as an electron donor to cytochrome oxidase

C/Superfamily: T. aquaticus cytochrome c552; cytochrome c6 homology

C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; pyroglyutamic a

F:2-87/Domain: cytochrome c6 homology <CY6>

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:11,14/Binding site: heme (Cys) (covalent) #status experimental

F:15,69/Binding site: heme iron (His, Met) (axial ligands) #status experimental

Query Match 6.1%; Score 161; DB 1; Length 131;

Best Local Similarity 37.3%; Pred. No. 0.00042;

Matches 44; Conservative 23; Mismatches 39; Indels 12; Gaps 6;

QY 396 KAGKATYDNCACHOPDGGVGNAPPLAN--SDYLNADHARASIVA--NGLSGKITV 451

Db 1 QADGAKIYAOCACGCHQONGGIGIPGAPPLAGHVAEILAKGGRBYLLVLYGIGQGIQEV 60

QY 452 NGNOYBSVMPALALSDQOIANVITYTLNSFGN-----KGGQ-LSADLVAK--AKKTP 501

Db 61 KGMKYNQWSSFPQLKDEIAPVILHATATAMGDAKKYKGFPTAEVKKLRKKLTP 118

RESULT 30

C82845

copper resistance protein A precursor XF0132 [imported] - Xylella fastidiosa (strain 9a5

C/Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000

C/Accession: C82845

C/Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequat

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; PMID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: C82845

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-611 <SIM>

A:Cross-references: GB:AE003866; GB:AE003849; NID:G9104906; PIDN:AAF82945.1; GSPDB:GN001

A:Experimental source: strain 9a5

R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincanti, A.P.; Ferreira, A.J.S.

Submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From

J.D.; Unquerra, M.L.; Kemper, E.L.; Kitzajme, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeira, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silvair

M.; Teunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C/Genetics:

A:Gene: XF0132

C/Superfamily: laccase

Query Match 6.1%; Score 159.5; DB 2; Length 611;

Best Local Similarity 20.1%; Pred. No. 0.004;

Matches 95; Conservative 57; Mismatches 136; Indels 185; Gaps 22;

QY 59 LPVIDAIVTHA--PEVPPVVDHDPKAVVYKMETEKVRLADGVEYQFWTFGGVPGOM 116

Db 39 LPV-----HAAPLPLPQORTTLELRIGSSVNFGRKRAI-----TVNLSLPGV 87

QY 117 IFRREGDTIEVQFSNH-PDSKMPHNVDFPAATGPG-----GAESAFTAPGHTSTFSFKA 170

Db 88 LEMRGDPTVQIHVNTLPD--VMTSIHMHGIVLPENMDGVPMSPDGIAFGHYLYRFQL 145

QY 171 LOPGLVYTHCAVAPVGMHIANMGVGLVLEPEGLP-KYDKYVYMGDFYKKGKYGEG 229

Db 146 HOSGTYWYH---SHMPFOAGLYGALIIDLEPPRYADREHILFSDW-----TD 194

QY 230 LOPFMEKXIRBDAYVFNVSGLATGENAKAKVGEVTFVNGGPNLTSFHVIGE 289

Db 195 LDPALFRKLKMSST-----DNTYQ-----RTVRF-----FHD1-- 225

QY 290 IDKVFEGEGKENHIQTL-----IPAG----- 314

Db 226 -----RRDGLRTLLADRMGMGRMTPTDLSVNAHYTYLLNGTTPAGAWTG 273

QY 315 -----GAITEPKVDVPGDYVLDVAIFRAFNKALGILKVEEENHEIYS 360

Db 274 LRPGEKVLRLINGSAMTVFVIRPG-----LKLTVAVDQ-----YV 313

QY 361 HQGT-----DANYLPEGAPO-AIDTQBAKT-----P 386

Db 314 HPVYDELRIALAEYVDVLIOPHGODAPAIIPADMGRTGYACGTLAVRGILAPLALDP 373

QY 387 APANLOEQIKKAGKATYDNCACHOPDGGVGNAPPLANSDYLNADHARAS 439

Db 374 RRLTMDQMGHMAHGQHSRASVMD-----MPCMTHPRL-GDHASHAS 418

RESULT 31

F70813

hypothetical protein RY0846c - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C/Accession: F70813

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S

C/Conor, R.; Davies, R.; Devlin, K.; Feltham, T.; Genies, S.; Hamlin, N.; Holtroyd, S

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:98295987; PMID:9634230

A:Accession: F70813

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1504 <COL>

A:Cross-references: GB:AL022004; GB:AL123456; NID:G3261550; PIDN:CAI17652.1; PID:G29169C

A:Experimental source: strain H37RV

C/Genetics:

Nature 406, 959-964, 2000
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A/Reference number: AB2950; MUID:20437337; PMID:10984043
A/Accession: AB3363
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-439 <STO>
A/Cross-references: GB:AE004657; GB:AE004091; NID:9948287; PIDN:AA05654.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA2266
C/Superfamily: membrane-bound alcohol dehydrogenase cytochrome c; cytochrome c6 homology
C/Keywords: chromoprotein; heme; iron; metalloprotein
F/40/Binding site: heme (Cys) (covalent) #status predicted
F/44/Binding site: heme iron (His) (axial ligand) #status predicted
F/188/191/Binding site: heme (Cys) (covalent) #status predicted
F/192/Binding site: heme iron (His) (axial ligand) #status predicted
F/324/327/Binding site: heme (Cys) (covalent) #status predicted
F/328/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match 5.8%; Score 153; DB 2; Length 439;
Best Local Similarity 35.5%; Pred. No. 0.0071;
Matches 38; Conservative 15; Mismatches 48; Indels 6; Gaps 4;

Db 396 KAGKATYDSNCAACHOPDGKGVNPAFPPLA-NSDYLNADHARAASIYANGLSGKITVNGN 454
314 KTGALVYVNDGCGCHRDTGKGYARVPALAGNPTVGSPTSLSVHVLKG-GLTPATHQ 371

Qy 455 QYES-VMPALA--LSDQIANVITYTLNFGNKGQGLSADVAKAKK 498
372 APSSTFMPFGWMNDDEIDAVVNFRTSWGNGQSPSVDEVRLRK 418

RESULT 35
T04343
L-ascorbate oxidase (EC 1.10.3.3) - rice (fragment)
C/Species: *Oryza sativa* (rice)
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C/Accession: T04343
R/Tabata, K.; Hirose, A.; Saaka, M.
Submitted to the EMBL Data Library, June 1997
A/Description: Gene expression and function of ascorbate oxidase in rice.
A/Reference number: 215298
A/Accession: T04343
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-380 <TAB>
A/Cross-references: EMBL:AB004799; PIDN:BA20520.1
A/Experimental source: subsp. *Gramineae*
C/Superfamily: laccase
C/Keywords: oxidoreductase
F/21-237/Disulfide bonds: #status predicted

Query Match 5.8%; Score 151.5; DB 2; Length 380;
Best Local Similarity 22.4%; Pred. No. 0.0075;
Matches 88; Conservative 47; Mismatches 138; Indels 119; Gaps 17;

Qy 101 VERQFWT-----FGQVPGOMTRVREGDTIEVQFSN--HPDSKPHNVDFNAATG 148
12 VERVLMAPDCQORVMGINGRFRPNITARAQVISTVMNNKMTGCVIHMGIROFGT 71
Db 149 PGGAASFTA-----PGHTSFSPKALOPGLVYVHCANAVPCHMIANGMYGLIVERKEG 204
72 PMADGTASISOCANVPSTFYKFAVDKPGTYFNH---GHFGMRAGLIGSLVLDSPR 128

Qy 205 LPKVDKEYV-----VMQDFTYTKGCGE-QGLAQFDEMEKAIREDAEVVFNG----- 250
129 QPEFRHQYDDGSELPMMLISDMHNVVAAAGLKDKNHFMETIGRPQITLLNGRQFE 188

Qy 251 -SYG-----ALTGENALAKV-----GE 267
Db 189 CTLAGPARKSEFKLNGERDPASTRRCATRRSAREERVGRCPRSGCAPVVFVNGGK 248

Qy 268 TVRLFVNGGPNLTSSFHVIGELIPDKVAFEGGKGENHNIQTLLPAGAAITTEKVDVPG 327
249 TYRLALIAS-----TSLSL-----NVKIQGNK-----MTVADGNHVEPFVDDID 291
Db 328 DYVLVDHAI-FPAFKGALGILKVEGENH-----EYSHKQTD-----VYLPFGAPQ 375
292 IYSGESYSVFFKADCKPASYWISVGRSHRPKTPALAILSYGNMAAPPLQFPAGEP- 350

Qy 376 AIDTQEPKTPAPALOE-----QIKAGKAT 401
351 -----PVTPAMNDTORSKAFYTSIRAKDT 375

RESULT 36
F83631
cytochrome c oxidase, subunit II PA0105 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C/Species: *Pseudomonas aeruginosa*
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: F83631
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lam,
J.; Lozy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A/Reference number: AB2950; MUID:20437337; PMID:10984043
A/Accession: F83075
A/Status: preliminary

Query Match 5.7%; Score 150; DB 2; Length 374;
Best Local Similarity 22.0%; Pred. No. 0.0093;
Matches 63; Conservative 38; Mismatches 86; Indels 100; Gaps 11;

Qy 265 VGETVRLFVNGGPNLTSSFHVIGELIPDKVAFEGGKGENHNIQT--TLIPAGA----- 316
135 LQGDVEYFSNLTPO-----DOINRQAKDEHYLVEVPLVPGTKVRLI 182
Db 317 -----ATTEF-----KYDVBQDY-----VLVDHA 335
183 TSSDVIHSMWVPAFAVKDPAIFGVNEAMTKVDEIGIYRQCAELCGKHGFMPIVVD-- 240

Qy 336 IFPAFNKALGILKVEGENHEIYSHKQTDVYLPFGAPQALDTPKTPAPALQEOI 395
241 -----VKPKAEFDQWLAKRKE-EAKKVE-----LTSKWTX-----EBLV 275

Qy 396 KAGKATYDSNCAACHOPDGKGVNPAFPPLANSDYLNADHARAASIYANGLSGKITVNGN 455
276 ARGDKVYHTTCAACHQABGQGMPEMPALKSGKITVGPENHLEVVNGVPG----- 327

Qy 456 YESVMPALA--LSDQIANVITYTLNFGNKGQGLSADVAKAKTK 500
328 --TAAFAFGKQNLNEVDLAIVITERNANWGNDDGWTPRQDVAAVYQK 372

RESULT 37
E83075
probable cytochrome c PA4571 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C/Species: *Pseudomonas aeruginosa*
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: E83075
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lam,
J.; Lozy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A/Reference number: AB2950; MUID:20437337; PMID:10984043
A/Accession: E83075
A/Status: preliminary

A.Molecule type: DNA
A.Residues: 1-675 <STO>
A.Cross-references: GB:AE004870; GB:AE004091; NID:g9950807; PIDN:AA07959.1; GSPDB:GN001
A.Experimental source: strain PA01
C.Genetics:
A.Gene: PA4571

Query Match 5.7%; Score 150; DB 2; Length 675;
Best Local Similarity 26.3%; Pred. No. 0.02;
Matches 46; Conservative 28; Mismatches 73; Indels 28; Gaps 6;

QY ILKVGGENHEIYSHKQTDVY-----LPGAPPAIDTOEAPKTP 386
DB 244 LKLT-GQNRFAIVSGSMTDVHHSIQHFSDDLLAIAYSIKSLPAGKDDLPMPSERRPLA 302
QY 387 APANTOEQIKAGKATYDSNCAACHOPDGKGVNAPPPILANSYDLNADHARAASIVANGLS 446
DB 303 APVDLYS--SRGGLGYAQCSCDKKDSGSGVGMFPPLGNPTVAS--AMFSTLHITLT 358
QY 447 GKIVNGQYESV--MPALA-LSDOQIANVITYTILNSFGNKGQSLADDVAKAK 498
DB 359 GMKTAQTATSHRVYTMPGFAQLEDEIRIETILSFVSSWGNSGSIIDAGGVKKLRQ 413

RESULT 38
F83387
copper resistance protein A precursor PA2065 [imported] - Pseudomonas aeruginosa (strain C)Species: Pseudomonas aeruginosa
C.Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #ext_change 31-Dec-2000
C.Accession: F83387
R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, N.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A.Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A.Reference number: A82950; MUID:20437337; PMID:10984043
A.Accession: F83387
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-632 <STO>
A.Cross-references: GB:AE004633; GB:AE004091; NID:g9948070; PIDN:AA05453.1; GSPDB:GN001
A.Experimental source: strain PA01
C.Genetics:
A.Gene: pcoa; PA2065
C.Superfamily: laccase

Query Match 5.7%; Score 149; DB 2; Length 632;
Best Local Similarity 21.3%; Pred. No. 0.022;
Matches 87; Conservative 59; Mismatches 149; Indels 114; Gaps 21;

QY 107 TFGGVPGOMIRVREGDTREVOFSNH--PDSK-----MPHNVDFHATGPGGAEASF 157
DB 66 TINSGLPPTLRMRGSDVTLRVRRLAEDTSHHGGIILPANMD-----GVPGLSF 117
QY 158 --TAFGHTSTSFKAQLQGLVYVHCVAVPVGNHIANGVGLILVEPKGLP-KVDKEYVV 214
DB 118 EGIAPGGIYEVRFKVRQNGTYWYH--SHSGIQEQAGVYALVIDAREPEPSYRDYV 174
QY 215 MCGPFYTGKGTGEGQIQPFDEKAKIREDAEYVVF-----NGSIALGTEN- 259
DB 175 LLTWSDE-----KQRIILAKLKQSDYVNFHRTVGDFTIDVSNAGMATLADRM 226
QY 260 --ALK-----AKVGETVRLFVNGGSP--NLTSFHYIGEIFDKVHFGGKGENHNIQT 308
DB 227 WAEMKMSPTDLADVSGYTYTTLINGQPPDGNTGIFR-----PGEKRLRF 272
QY 309 TLIPAGAAITEFKVDVG-----DYVLVDH--AIFRAFNKALGILKVEG 352
DB 273 ----VNASAMSYFVRIQGLKMTVAADQGHVEPVSVBELRIAVAEYTDV-----IVEPQG 324
QY 353 EENHEIYSHKQTDVAVYLBEGAPQALDTQEAR-KTAPANLQEQIRAKATYDSNC--AA 408
DB 325 ERAYTLFAQMSDRSGY-ARGTLALAEGLSAPVPTDPBPPLICMDDMGGMGMDHGMGHGA 383

QY 409 CHOP-----DGKVPNAPPPILANSYDLNADHARAASIVANGL 445
DB 384 ARPRPASEMDHSMKSGMDVTCGMDS--KXAGMDSGMDSKXAGMDSNMGM 430

RESULT 39
S14271
membrane-bound alcohol dehydrogenase (EC 1.1.1.-) cytochrome c precursor - Acetobacter I
N.Alternate names: membrane-bound alcohol dehydrogenase 44k chain
C.Species: Acetobacter polyoxogenes
A.Variety: strain NB11028
C.Date: 21-Nov-1993 #sequence_revision 09-Aug-1996 #ext_change 17-Nov-2000
C.Accession: S14271
R.Tamaki, T.; Fukaya, M.; Takemura, H.; Toyama, K.; Okumura, H.; Kawamura, Y.; Nishiyama
Biochim. Biophys. Acta 1088, 292-300, 1991
A.Title: Cloning and sequencing of the gene cluster encoding two subunits of membrane-bc
A.Reference number: S14270; MUID:91159482; PMID:2001402
A.Accession: S14271
A.Molecule type: DNA
A.Residues: 1-468 <TAM>
A.Cross-references: GB:D00635; NID:G216185; PIDN:BA00529.1; PID:G216187
A.Experimental source: strain NB11028
C.Complex: heterodimer of 72k and 44k (cytochrome c) chain
C.Superfamily: membrane-bound alcohol dehydrogenase cytochrome c; cytochrome c6 homology
C.Keywords: alcohol metabolism; blocked amino end; chromoprotein; electron transfer; hen
F.1-23/Domain: signal sequence #status predicted <SIG>
F.24-468/Product: membrane-bound alcohol dehydrogenase cytochrome c chain #status predic
F.320-403/Domain: cytochrome c6 homology <Cys>
F.45/48/Binding site: heme (Cys) (covalent) #status predicted
F.149/Binding site: heme iron (His) (axial ligand) #status predicted
F.193/196/Binding site: heme (Cys) (covalent) #status predicted
F.197/Binding site: heme iron (His) (axial ligand) #status predicted
F.330/333/Binding site: heme (Cys) (covalent) #status predicted
F.334/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match 5.6%; Score 146; DB 1; Length 468;
Best Local Similarity 25.2%; Pred. No. 0.023;
Matches 84; Conservative 50; Mismatches 127; Indels 72; Gaps 19;

QY 184 PYGNHIANGVGLIL-VEPKGLPVYDKEYVYMGDFYTKKYGQGL---QPFDM-EK 237
DB 149 PLSMKRWPLGIWRMSPSPKCDFTPAFGDPETARDYLVGTG-GHCGAHTPRGPMOER 207
QY 238 AIREDAEYVFNKSGALITGENALAKAYGETYR-----LFVNGGNNLTSFHVIEIFPK 293
DB 208 AL--DAA-----GGPFLSGGAPIDMWAPSLRNDPVVGLGWSBDDIYTFUKSGRIDHS 260
QY 294 VHEFGKGENHNIQTLLIPAGGAITEFKVDVPGDYLVLD--HAIFRAFNKALGILKVE 351
DB 261 AVF-GGMD-----VVMSTQYTFDDDLALAK-YLKSLLPVPPSQ 299
QY 352 EENHEIYSHKQTDVAVYLBEGAPQALDTQEARPPAPANLQEQIRAKATYDSNCAACHQ 411
DB 300 GN-----YTYDPSTANMLASG-----NTASVP-----GADTYKECAICHR 335
QY 412 PDGKVPNAPPPILANSYDLNADHARA-ASIVANGISGKITVNGNYESV-MPAL--ALSD 467
DB 336 NDGGVAVMPPLAENPVVVTENPTSLVNVLAHG--GVLPSNMAPSAVAMGYSLSLA 393
QY 468 QOIANVITYTILNSFGNKG-QOLSADVDVAKAKT 499
DB 394 QOIAADVNVFITSWGNKAPGVITADYKLAADT 426

RESULT 40
B49340
membrane-bound alcohol dehydrogenase (EC 1.1.1.-) cytochrome c precursor - Acetobacter f
N.Alternate names: membrane-bound alcohol dehydrogenase 44k chain
C.Species: Acetobacter pasteurianus
C.Date: 07-Apr-1994 #sequence_revision 09-Aug-1996 #ext_change 17-Nov-2000
C.Accession: B49340
R.Takemura, H.; Kondo, K.; Horinouchi, S.; Bepu, T.

J. Bacteriol. 175, 6857-6866, 1993
 A:Title: Induction by ethanol of alcohol dehydrogenase activity in Acetobacter pasteuriae
 A:Reference number: A49340; MUID:94042848; PMID:8226628
 A:Accession: B49340
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-472 <TAK>
 A:Cross-references: GB:D13893; NID:6517067; PIDN:BAA02993.1; PID:9452587
 A:Experimental source: strain NC11380
 C:Complex: heterodimer of 72K and 44K (cytochrome c) chains
 C:Superfamily: membrane-bound alcohol dehydrogenase cytochrome c; cytochrome c6 homology
 C:Keywords: alcohol metabolism; blocked amino end; chromoprotein; electron transfer; hem
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-472/Product: membrane-bound alcohol dehydrogenase cytochrome c chain #status predic
 F:321-404/Domain: cytochrome c6 homology <CY6>
 F:46/49/Binding site: heme (Cys) (covalent) #status predicted
 F:50/Binding site: heme iron (His) (axial ligand) #status predicted
 F:194/197/Binding site: heme (Cys) (covalent) #status predicted
 F:198/Binding site: heme iron (His) (axial ligand) #status predicted
 F:331/334/Binding site: heme (Cys) (covalent) #status predicted
 F:335/Binding site: heme iron (His) (axial ligand) #status predicted.

Query Match 5.5%; Score 145.5; DB 1; Length 472;
 Best Local Similarity 33.3%; Pred. No. 0.025;
 Matches 39; Conservative 18; Mismatches 31; Indels 29; Gaps 5;

Qy	397	AGKATYDSNCA	CHOPDGKGV	PNAPPL	NSDYLNADH--	ARASIVANGIS	SKITVNGND	455
Db	322	AGAKTYVEOCAI	CHRDGGGV	ARMFPL	AGNPVVS	DNPTSV	NAHIVDG-----	370
Qy	456	YESVMP-----	AIA-----	LSDOQIAN	VITYTLN	SPGNGK-	GOLSAD	VDVAKAK 497
Db	371	--GVLEPT	WAPSAV	AMPDYKNIL	SDQIAD	VNFIR	SAWGNRAP	ANTPADIQKLR 425

Search completed: August 27, 2003, 18:32:44
 Job time : 46 secs

QY 336 IFRPANKGALGILKEBENHEIYSHKOTDAVYLPEGAPOAIDTOEAPKTPAPANLOEQI 395
 Db 241 -----VKPKEFQOMLAKRKE-EAAKYKE-----LTSKEWTX-----ELV 275
 QY 396 KAGKATYDNCAACHQDQGVPAFPPLANSIDLNDHARAASIVANGLSGKITVGNQ 455
 Db 276 ARGDQYVHTICACHQAGQGMFPMFPALGSKITVTPGKHHLELVNGLVPG----- 327
 QY 456 YESVMPAIA--LSDQOIANVITYTLNSFGNKGQSLADDAVAKAKTK 500
 Db 328 --TAMAFGKQNEVDLAAYITVERNAMGNDGDMVTPKQVAVAKOK 372

RESULT 2
 US-09-732-350-5
 ; Sequence 5, Application US/09732350
 ; Patent No. US20010031490A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Svendsen, Allan
 ; APPLICANT: Xu, Feng
 ; TITLE OF INVENTION: LACCASE MUTANTS
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. US20010031490A10 No. US20010031490A1disk of No. US20010031490A1ch
 ; STREET: 405 Lexington Avenue
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10174
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/732.350
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/032,315
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rozek, Carol
 ; REGISTRATION NUMBER: 36,993
 ; REFERENCE/DOCKET NUMBER: 5200.200-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 529 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-732-350-5

Query Match 5.4%; Score 141; DB 9; Length 529;

Best Local Similarity 20.4%; Pred. No. 0.00026; Indels 126; Gaps 20;

Db 59 LPVIDAIVTHAPEVPPVDRDHPAKVVK--METVEKWLADGVEYQFTFGGVPGOM 116
 8 LPLLAAVST-----PAPAAVANYKFDIKNNVAVPDGFGRSIVSVNGVLPGL 54

QY 117 IRRREGDTLEVQFSNH---PDSKMPHNVDFH---AATGGGGAESAFTAP-----GHTS 164
 Db 55 ITANKGDTLRINVTQLTDPKMRATTIHHHGLFOATTADEDEGPAFTVQCPIAQNLSYTY 114

QY 165 TFSFKALQPLGLVYHCAVAPVGMHIANGMYG-LILVER---KEGLPKVDKEYVYVNGQDF 219
 Db 115 EIRPLRG-QGTGMWYH--AHLASQYVVDGLRGPLVYIDPNDPHKSRDYVDVDASTVWMLDWM 170

QY 220 Y-----TKGKXGEGQ---LOPFDMEKAIREDAEVYVNGSVGALTGSENAKAKVGETV 269
 Db 171 YHTPAPVLEKQMFSTNNNTALLSPVDSGLINGKGRVY-----GPAVPRSVINVKRGKRY 225
 QY 270 RLFVNGGNGNLTSFHVIGIEIDPKHFBEGKGENHIQTLIPAGS-----AATFEKVD 324
 Db 226 RLKVINASAIGSTFTSI-----EGHSL--TVLEADGILHQPPLAVDSFQIY 268
 QY 325 VPGDVLVVDHALFRANKGALGILKEBENHEIYSHKOTDAVYLPEGAPOAIDTOE--- 381
 Db 269 AGGRYVIVYEANQTAANYIRAPMTYAGAGTANLDPNTVFALVHLEGAPNAEPTTEQGS 328
 QY 382 -----APKTPAPANLOEQIKAGKATYDS-----NCAACHQDPDGKV 417
 Db 329 AIGTALVEENLHALINPGAPGSA PADVSLNLAIGRSTVDGILRFTFNNIKYEAP----- 383
 QY 418 PNAFPPLANSIDYINADHARAASIVANGLSGKITVANGNOYESVMPAIALSDQOIANVITYT 477
 Db 384 --SLPPL-----LKITLANNASNDADFTPNEHTIVLPNKVIELNITGADHP 428
 QY 478 LNSFGNKGQSLADDAVAKAKTKRN 502
 Db 429 IHLHGH-----VFDIVKSLGGTPN 447

RESULT 3
 US-09-869-877-5
 ; Sequence 5, Application US/09869877
 ; Publication No. US20020192792A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Schneider, Palle
 ; APPLICANT: Danielson, Steffen
 ; APPLICANT: Svendsen, Allan
 ; TITLE OF INVENTION: laccase Mutants
 ; FILE REFERENCE: 10179.204-US
 ; CURRENT APPLICATION NUMBER: US/09/869,877
 ; CURRENT FILING DATE: 2001-07-06
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 529
 ; TYPE: PRT
 ; ORGANISM: Rhizoctonia solani
 ; US-09-869-877-5

Query Match 5.4%; Score 141; DB 10; Length 529;

Best Local Similarity 20.4%; Pred. No. 0.00026; Indels 126; Gaps 20;

Db 59 LPVIDAIVTHAPEVPPVDRDHPAKVVK--METVEKWLADGVEYQFTFGGVPGOM 116
 8 LPLLAAVST-----PAPAAVANYKFDIKNNVAVPDGFGRSIVSVNGVLPGL 54

QY 117 IRRREGDTLEVQFSNH---PDSKMPHNVDFH---AATGGGGAESAFTAP-----GHTS 164
 Db 55 ITANKGDTLRINVTQLTDPKMRATTIHHHGLFOATTADEDEGPAFTVQCPIAQNLSYTY 114

QY 165 TFSFKALQPLGLVYHCAVAPVGMHIANGMYG-LILVER---KEGLPKVDKEYVYVNGQDF 219
 Db 115 EIRPLRG-QGTGMWYH--AHLASQYVVDGLRGPLVYIDPNDPHKSRDYVDVDASTVWMLDWM 170

QY 220 Y-----TKGKXGEGQ---LOPFDMEKAIREDAEVYVNGSVGALTGSENAKAKVGETV 269
 Db 171 YHTPAPVLEKQMFSTNNNTALLSPVDSGLINGKGRVY-----GPAVPRSVINVKRGKRY 225

QY 270 RLFVNGGNGNLTSFHVIGIEIDPKHFBEGKGENHIQTLIPAGS-----AATFEKVD 324
 Db 226 RLKVINASAIGSTFTSI-----EGHSL--TVLEADGILHQPPLAVDSFQIY 268

QY 325 VPGDVLVVDHALFRANKGALGILKEBENHEIYSHKOTDAVYLPEGAPOAIDTOE--- 381
 Db 269 AGGRYVIVYEANQTAANYIRAPMTYAGAGTANLDPNTVFALVHLEGAPNAEPTTEQGS 328

QY 382 -----AKTPAPANLEQIYKATYDS-----NCACHQDPQKGV 417
DB 329 AIGTALVEENLHALINPGAFGSGAPADVSINLAIGRSTVDGILRFTFNINRYEAP----- 383
QY 418 PNAFPLANSDYLNADHARAASIVANGLSKITVNOQVESVPALALSQOJANITYT 477
DB 384 --SLEPTL-----LKITLANNASNDADFTNEHTITVLPNNKIVLELNTGADHP 428
QY 478 LNSFGNGGOLSDADVAKAKTKPN 502
DB 429 IHLHGH-----VFELVKSIGCTPN 447
RESULT 4
US-09-738-626-6780
; Sequence 6780, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6780
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6780
Query Match 5.4%; Score 140.5; DB 10; Length 493;
Best Local Similarity 19.0%; Pred. No. 0.00026;
Matches 80; Conservative 57; Mismatches 174; Indels 109; Gaps 15;
QY 15 LSLMLSGSNQADKAAQPKSSTVDAAKTANMDNNAQDHGELFVIDAIVTHAEVPP 74
DB 12 LGLGLVAG-----TGAVAACTSDPGPAAS-----APG 38
QY 75 PVDURDPAKVVVMEETVEKVMRL-----ADGVEYQFTFGGQVPGCMIRVREGDTIEVO 128
DB 39 PSRLPRTFTALCEPFTVRRITLTPRLSLDIOGIEATKGVSTGDAIEATIGDVLQVD 98
QY 129 FSNH-PDSKMP--HNVDFAATGPGGAAEASFTAPGHTSTFSFKALQPLVYVHCAVAP 185
DB 99 ITNELPESTSIIWHGJIALHNAADGVPGMTQDPIEPGESFSYVEVPHGTYFYH---SHT 155
QY 186 GMIHANGMGLILVERKEGLPKVDKEYVMQSGPYTKGKKGEGLOPDMEK----- 237
DB 156 GLQDLRGHLAPLIRPDQAEQDDVETWTVLDDWVD---GIQGTPEDELDELKLTGMSGD 211
QY 238 -----AIRBDAEYVFNQSVGAL-----TGENALKAKVGETVRL-FV 273
DB 212 HNGRMGMGHHGMMHGTTPRVLGSDGVVWVPHYLNGRIPRNHRTFEPARPEDKALRRT 271
QY 274 GNGGPNLTSSPHVI--GEIFDKYHFGGKGENHNIOITLLIPAGGAALTEPKVDV--GD 328
DB 272 NSGGDTI---FKVALGCHMTVTHTDGFPVQWETESIYLSMGE-----RVDVEVILGD 322

QY 329 YVLVDHAFRANFGALGILKVEGENHE-----IYSHKQTDVAVLEPGAP 374
DB 323 GIEPLTALAVGDDRAFAVIRTAGGQAPRVDVDFPELSSTGLLSLTKRADRLLEGRP 382
RESULT 5
US-10-174-693-291
; Sequence 291, Application US/10174693
; Publication No. US20030131373A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.100305
; CURRENT APPLICATION NUMBER: US/10/174, 693
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 291
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-10-174-693-291
Query Match 5.2%; Score 136.5; DB 16; Length 326;
Best Local Similarity 20.9%; Pred. No. 0.00032;
Matches 71; Conservative 46; Mismatches 129; Indels 93; Gaps 14;
QY 78 RDHPAKVVVMEETVEKVMRLADVEQVFTFGGQVPGCMIRVREGDTIEVPSNHDSDK 137
DB 9 RNYTFNVVKNKTT-----RLCS--SKPIYVAGMPFGPTLYABEDTVLVRVSN---RV 57
QY 138 PHNVDFH-----AATGPGGAAEASFTAPGHTSTFSFKAL-QPGLYVYHCAVAPG 186
DB 58 KYNVTIHHMGIROLKRTGMAQDGPAYITQCP--QPGSYVYNFTITGGRGLTLMHAIH---- 112
QY 187 KHIANGMGLILVERKEGL---PVVDKEYVMQSDPYTKGKYG-----EGLOPDMK 236
DB 113 LMLRATLHGAIYILPRGVPPFPKPKKEVVVVLGSMWKSDEGVISQAISGLAP----- 168
QY 237 KAIREDAEYVNVGSGALT-----GENALKAKVGETVRLFPVNGGPNLTSSPHVIGELF 291
DB 169 ---NVSDHTITNGHPGSSNCPGSGFTLPVSESGKYLRIINALNELFFKLAGH--- 222
QY 292 DKVHFGGKGENHNIOITLLIPAGGAALTEPKVDVPGDVLVDHAFRANFGALGILKVE 351
DB 223 -----QITVEVDATVVPFKTDT-----IVIAP 246
QY 352 GEENHEIYSHKQTDVAVLEPG-----APOAIDTOEAPKT 385
DB 247 GQTTNALISTDQSSGKVVVAASPFDSPFAVDNMTATAT 285
RESULT 6
US-09-732-350-8
; Sequence 8, Application US/09732350
; Patent No. US20010031490A1
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Xu, Feng
; TITLE OF INVENTION: LACCASE MUTANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4553
LENGTH: 511
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4553

Query Match 5.0%; Score 131.5; DB 10; Length 511;
Best Local Similarity 22.5%; Pred. No. 0.0015;
Matches 104; Conservative 51; Mismatches 199; Indels 109; Gaps 21;

29 KAAQKSTVYDAAKAANNDNAASOEHOGE--LPVIDAIVTHAPVPPVD--RDHPAK 83
11 KAGVLAATVVGNOUVVACSSSDVRGEGEPRLPY-----PRADGTREGSS 58
84 VVVKETVEKVRKLADVEYQVFTFGOVPGVMIRREGDTLEVOFSNHPDS----- 135
59 VHPALAEQGESQILPDDVTTKTGFGNGTHLGPTLVVKKGDVAVDINNIDEMTIVHWG 118
136 -KMPHNVD--HAATPGGGAESFTAPGHTSTFSKALQPGLYVYHCAYVGMHIANG 192
119 MTLPALADGPHSPICPGQTWSPMTVANDNATLWHPHNGI-----TGLHAYRG 169
193 MYGLILVERKEGLPKYD--KEYVV-----MGSDYTKGKVGEOGLQFPDMKAIRED 243
170 LKGMITVE--DEATDQKDLPERGVDDIPLVIMDRHLEDSLDEELP--DLGLDGTPT 226
244 EYVVENGVGALTGENALAKAVGETVRLFVNGGPNL-----TSSFFVIG-----EIPD 292
227 ANGITAHFDATTRRVRFRVLNNSMRFY-----NLAFSDRTFQVIASDSGLDEPQD 280
293 KHFBSGCKENNNIQTLLIPAGGAIT-----EFKVDVPGDYVLVDHAI FRAFNKALG 346
281 RTLLAIGPERWEIIVLELP--GEDVTLESVGFEDNVGVDPDEFVDFGMSDSFQ----- 333
347 LKVESENHETIYSHKQTDAAVLP-----EGAPQAITDQAP--KTPRANLQEOI 395
334 LLLITGPSP-----DAAQAPALPGVLVSTEDVIDATERITIMTFISINDLOMDM 384
396 KAGKATYDSNCAACHQPDGKGVNAPFPPLANSDYLNADHARAA 438
385 QAVDVVIDHD-----QPEWIVITN-----DNSDMENFHVHDA 417

RESULT 9
US-10-115-563-14
Sequence 14, Application US/10115563
Publication No. US20030008307A1
GENERAL INFORMATION:
APPLICANT: Griffin, John H
Greengard, Judith S
TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
AND COMPOSITIONS THEREOF
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of

Patent Counsel
STREET: 10666 No. US20030008307A1ch Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/115,563
FILING DATE: 02-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/410,488
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Filting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: 449.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2224 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-115-563-14

Query Match 4.9%; Score 128.5; DB 15; Length 2224;
Best Local Similarity 19.7%; Pred. No. 0.036;
Matches 82; Conservative 61; Mismatches 150; Indels 123; Gaps 20;

63 DAIVTHAPVPPVRDRHPAKVVMKMT-----VEKVMKLADGVEYQVFTFGQ----- 111
358 EYIMDYAPVPIPMNDKYSQHLNFSNOIGHYKKVM-----TQYEDSFTKHTVPM 413
112 ---VPGMIRVREGDTLEVOFSN--HDSKMPHNVDFAATPGGGAESFTAPGHTS 164
414 KEDGLIGPITRAQVBDTLKIVFKMASRYSIYPHGVTSPYEDE--VNSPTS-GRNN 469
165 TFSFKALPG-LYV-----HCAVAP-----VGMHIANGMYGLIIVERK 202
470 TM-IRAVQGETYTYKWNILFEDDEPTENDAOCLTRPYSDVDIMRDIAGLIGLILICKS 528
203 EGLPK-----VDKEYVVMQGF-YTKGKYGEOGLQFP--DMBKAIREDAEVVFNGSVG 253
529 RSLDRGIRPADIEQAVFAVFDENKSWYLEDNINKFENDEYKRDPPK----- 580
254 ALTGENALAKAVGETVRLFVNGGPNLTSFFVIGETIPD-----KVAF 296
581 -----YESNMSTINGV-----PESITTLGFCFDDYQVMHFCVGTQNEILTIHF 626
297 EGK--GENNNIQTLLIPAGGAITEFKVDVPGDYVLVDHAI FRAFNKALGILKVESE 353
627 TGHSPFYGRHEDTLTFPMRGESVT-----VTMD-----NVGTWMLTSMNS 669
354 ENHEIYSHKQTDAAVLPBEGAQAITDQAPKTPAP-----NLOBQIKAGKATYD 403
670 PRSKLRLEKFDVKCIPDDEDSYEIFEPPBSTVWATRKMDRLPEDESDADVD 725

RESULT 10
US-10-172-712-31
Sequence 31, Application US/10172712
Publication No. US20030125232A1
GENERAL INFORMATION:
APPLICANT: GRIFFIN, JOHN H.

APPLICANT: GALE, ANDREW J.
APPLICANT: GETZOFF, ELIZABETH D.
APPLICANT: PELLEGUER, JEAN-LUC
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
FILE REFERENCE: 4198-4001US1
CURRENT APPLICATION NUMBER: US/10/172.712
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: 60/298,578
PRIOR FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 31
LENGTH: 2224
TYPE: PRT
ORGANISM: Homo sapiens
US-10-172-712-31

Query Match 4.9%; Score 128.5; DB 15; Length 2224;
Best Local Similarity 19.7%; Pred. No. 0.036;
Matches 82; Conservative 61; Mismatches 150; Indels 123; Gaps 20;

QY 63 DAIVTHAEVPPVDRDHPAKVNVKMET-----VEKWRLADGVEYQFTFGGQ----- 111
DB 358 EVIWDVAPVPIRANMDKRTSGHLDNFSNQIGKHKVMY-----TQYDESFTHKTVNPM 413
QY 112 ----VPGQMRVREGDTIEVOFSN--HPDSKMPHNVDFAATGPGGAASTFAPGHTS 164
DB 414 KEDGIIIGPIIRIAQVRDTLKIIVKMAARFYSIYPHGVTFSPYEDE--VNSSFYS-GRNN 469
QY 165 TFSFKALQF-LVYV-----HCAVAP-----VGMHANGYGLILVEPK 202
DB 470 TM-IRAVQGETYTKWNILFEDEPTENDACLTFRPYSDVDIMRDAISGLILGLCKS 528
QY 203 EGIPEK-----VPEKEYVMQGF-YTKGKGEOQLQGF--DMEKALREDAEYVFNVSVG 253
DB 529 RSLDRGIQAAIDIEQAVAVFEDENKSWLEDNINFCENPDEYKRDKEF----- 560
QY 254 ALTGENALKAKVGETVRLFYNGGPNLTSSPHYIGELFD-----KVHF 296
DB 581 -----YESNINSTINGV-----PESITTLGFCFDDTVQWFCVGTQNEILTIHF 626
QY 297 EGCG--GENHNITOTLIPAGGAITEFKVDVEDYLVVHAIFRANKALGLIKVEGE 353
DB 627 TGHSTFYKGHEDTLTFPMRGESVT-----VTMD-----NVSTWMLTSMSS 669
QY 354 ENHEIYSHKQTDVAVLPEGAPOAIDTQEAPEKTPAPA-----NLQEOIKKAKATYD 403
DB 670 PRSKRLKFRDVKCIPIDDDEDSYEIFEPPESTVMAIRKXHDRLEPDESDADYD 725

RESULT 11
US-09-732-350-7
Sequence 7, Application US/09732350
Patent No. US20010031490A1
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Xu, Feng
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20010031490A10 No. US20010031490A1disk of No. US20010031490A1ch
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/732.350

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/032,315
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5200.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-732-350-7

Query Match 4.7%; Score 124; DB 9; Length 572;
Best Local Similarity 21.4%; Pred. No. 0.011;
Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;

QY 86 VKMETVEKYMRLA-----DGVEYQFTFGGQVPGQMRVREGDTIEVOFSN--HPDSK 136
DB 17 VLARTVEYMLKISNGKIAPDGVBRDALTIVNGGYPGLIPANKGDTLKVQNKLTNPDMY 76
QY 137 MPHNVDFAA-----TGGGGAASFTAPGHTSFSFK-ALQGLYVYHCAVAPVGM 187
DB 77 RTTSHIMHDLQHRNADDDGPAFVTCPI-VQASATYTMPLGDDGTGYWYH--SHLS 132
QY 188 HIANGVY-LILVEPEKGPVKV---DKEYVMQGFY-TKGK---YGEQGLQPPDMK 237
DB 133 QYVDGRLGFLVYDPPKRLLYDIDDEKTVLLIGDMYHTSSKALATGNTLQOPDS-- 190
QY 238 AIREDAEYVFNVSVALTGEN-----ALKAVGETVRLFYNGGPNLTSSPHYIG 288
DB 191 -----ATNGK-GRPPDMTPANPNTLYTLKVRGKRYRLRVINS--SALASF----- 235
QY 289 ELFDKVFHFGCGENHNITOTLIPAGGAITEFKVDVEDYLVVHAIFRANKALGLIL 348
DB 236 ---RMSIOGHK-----MTVIADGVSTKPYQVD----- 260
QY 349 KVEGENHEIYSHKQTDVAVLPEGAPOAIDTQEAPEKTPAPA-----PAP-ANLQEOIKKAKATYD 404
DB 261 -----SPDILAGRIDAV-----VEANOEPDTYWINAPDLTVANKTAQALLIYED 305

RESULT 12
US-09-869-877-7
Sequence 7, Application US/09869877
Publication No. US20020192792A1
GENERAL INFORMATION:
APPLICANT: Schneider, Palle
APPLICANT: Svendsen, Allan
TITLE OF INVENTION: Laccase Mutants
FILE REFERENCE: 10179.204-US
CURRENT APPLICATION NUMBER: US/09/869,877
CURRENT FILING DATE: 2001-07-06
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 572
TYPE: PRT
ORGANISM: Rhizoctonia solani
US-09-869-877-7

Query Match 4.4%; Score 114.5; DB 9; Length 594;
Best Local Similarity 19.0%; Pred. No. 0.092;
Matches 123; Conservative 72; Mismatches 226; Indels 225; Gaps 28;

Query 6 LKTTLLICALSALMLSGCSNQADKAAQPKSIVDAAKT---ANADNAASQEHQGLPVI 62
Db 2 LFKSMQLAASGLISGLVIGIPMDTGSHP-IEAVDPEVKTEVPADSLILAAAGDDWESPYPY 60
63 DALVTAAPEVPP-----PV-DRD---HPAKVVVMKETEYKMRADGVEYQFWTF 108
Db 61 NLLYRNALPIPPVKOPKMIITNPVTGKDIWYEIEIKPFOQRIYPTLRPTLVGYD---- 116
109 GGVPCQMIRVREGDTIEVQFSNHPDSKPHNVDFHA--ATPGGGAASFTAPGHTSTF 166
Db 117 -GMSPEPTFNVRGTEVTVRFIN--NATVENSYHLGSPSRAPFDGMADVTFFPGEXKYD 173
Query 167 SFKALQPG--LYVYHCAVAPVGMHIANGYGLIV---EPKELPKVDKEYVYVMQGDY 220
Db 174 YFPNYQSARLLWYHDAFMKTAENAYFGAGAYIINDEADALGLP----- 219
Query 221 TKRGYEGGLOPDEKAIREDAEYVFNVSVALTGENA----- 260
Db 220 --SGYGE---FDIPLIL--TAKYVNADGTLRSTEGEDDLMGDVIVHNGQWPFLNVQ 270
Query 261 -----LKAKYGETVRLF--VNGGPNLTSFPHVIGE----- 289
Db 271 PKRYRFRFLNAVSRAMLVLYVTRTSSPNVRIFFOVIASDAGLLQAPVQTSNLYLAVER 330
Query 290 --IFDKVHEGKGGENNIQT-----LIPAG----- 314
Db 331 EIIIDFTNFAQTLLDRNVAETNDVGDEDEYARTLEVNRFPVSSGTVEDNSQVPSLTRDV 390
Query 315 -----GAATFEKVDVPGDVVLDVDAIF-----RAFKGALGILKYGEEN-----H 356
Db 391 PPPHKEGPADKHKFKERSNGHYLLINDVGFADVNERVLAKPELGIVAEVWELNSSGGWSH 450

US-09-338-723A-2
; ORGANISM: Stachybotrys chartarum
; LENGTH: 594
; TYPE: PRF
; ORGANISM: Stachybotrys chartarum
US-09-338-723A-2

Query 269 VRLFVNGGPNLTSFPHVIGEIPDKVHEGKGGENNIQTLLIPAGAAITFEKVDVPGD 328
Db 215 MRLISLSCDBNM--QFSIDGHELIIEVDGQLTEPHVLDLQIFTGRRYSFVLDANP-- 270
Query 329 YVLVDHAIFPAF--NKALGILKYGEENHEIYSHKQTDVAYLPEGAPOAIDTOEAPKTPA 387
Db 271 ---VDVWIRAPQPKGRNGLAGT-----FANGVSAILRYAGANADPTTSA--NPN 317
Query 388 PAULQEOIKAKGKATYDSNCAACHQPDQGVNAPPLANSDYLANADHARAASIVANGLSG 447
Db 318 PAQLNE-----ADLHALIDPAAPGIFP--FGAAD--VMLRQLOQPS-----GG 356
Query 448 KITVNGNQYESVMPAI-----ALSDQOIAN 472
Db 357 RFTINGTAYES--PSVPTLLQIMSGAQSAN 384

RESULT 15
US-09-338-723A-2
; Sequence 2, Application US/09338723A
; Patent No. US20020019038A1
; GENERAL INFORMATION:
; APPLICANT: Huaming, Wang
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC561-2
; CURRENT APPLICATION NUMBER: US/09/338,723A
; CURRENT FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 09/220,871
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 594
; TYPE: PRF
; ORGANISM: Stachybotrys chartarum
US-09-338-723A-2

Query 357 EYSH-----KOT-----DAVYLPEGAPOAIDTOEAPKTPA---N 390
Db 451 PVHILVDPKILKRTGROGWMPYBSAGLKVWLGRLGELTLEAHYQWPTGAYMHCHN 510
Query 391 LOEQIRAGKATYDSNCAACHQ-----PDQGVNAPPLANSDYLANAD--HAR 436
Db 511 LIHEDNDMAVAF--NVTAMEEKGYLQDEDEDPNNPKRAVP-----YVRNDFHAR 558
Query 437 AASIVANGLSKITVNGNQYESVMPAIALSDQOIANVITYTLNSFG 482
Db 559 AGNFSAESITARVQ-----ELAEQEPYRLDEILIEDLG 591

RESULT 16
US-10-080-210-2
; Sequence 2, Application US/10080210
; Patent No. US20020142423A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Huaming
; APPLICANT: Bodie, Elizabeth A.
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC561-3
; CURRENT APPLICATION NUMBER: US/10/080,210
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/220,871
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 09/338,723
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 594
; TYPE: PRF
; ORGANISM: Stachybotrys chartarum
US-10-080-210-2

Query Match 4.4%; Score 114.5; DB 14; Length 594;
Best Local Similarity 19.0%; Pred. No. 0.092;
Matches 123; Conservative 72; Mismatches 226; Indels 225; Gaps 28;

Query 6 LKTTLLICALSALMLSGCSNQADKAAQPKSIVDAAKT---ANADNAASQEHQGLPVI 62
Db 2 LFKSMQLAASGLISGLVIGIPMDTGSHP-IEAVDPEVKTEVPADSLILAAAGDDWESPYPY 60
63 DALVTAAPEVPP-----PV-DRD---HPAKVVVMKETEYKMRADGVEYQFWTF 108
Db 61 NLLYRNALPIPPVKOPKMIITNPVTGKDIWYEIEIKPFOQRIYPTLRPTLVGYD---- 116
109 GGVPCQMIRVREGDTIEVQFSNHPDSKPHNVDFHA--ATPGGGAASFTAPGHTSTF 166
Db 117 -GMSPEPTFNVRGTEVTVRFIN--NATVENSYHLGSPSRAPFDGMADVTFFPGEXKYD 173
Query 167 SFKALQPG--LYVYHCAVAPVGMHIANGYGLIV---EPKELPKVDKEYVYVMQGDY 220
Db 174 YFPNYQSARLLWYHDAFMKTAENAYFGAGAYIINDEADALGLP----- 219
Query 221 TKRGYEGGLOPDEKAIREDAEYVFNVSVALTGENA----- 260
Db 220 --SGYGE---FDIPLIL--TAKYVNADGTLRSTEGEDDLMGDVIVHNGQWPFLNVQ 270
Query 261 -----LKAKYGETVRLF--VNGGPNLTSFPHVIGE----- 289
Db 271 PKRYRFRFLNAVSRAMLVLYVTRTSSPNVRIFFOVIASDAGLLQAPVQTSNLYLAVER 330
Query 290 --IFDKVHEGKGGENNIQT-----LIPAG----- 314
Db 331 EIIIDFTNFAQTLLDRNVAETNDVGDEDEYARTLEVNRFPVSSGTVEDNSQVPSLTRDV 390
Query 315 -----GAATFEKVDVPGDVVLDVDAIF-----RAFKGALGILKYGEEN-----H 356
Db 391 PPPHKEGPADKHKFKERSNGHYLLINDVGFADVNERVLAKPELGIVAEVWELNSSGGWSH 450

QY 357 ELYSH-----KOT-----DAVYLBEGAPQAIIDTOEAKTPAPA---N 390
DB 451 PVHILVDEKIKRTRGQVMPYESAGLKDYYWLGREGTLLTEAHYQWTAIYMHCHN 510
QY 391 LOEOIKAGKATYDSNCAACHQ-----PDGKGVNAPFPLANSDYLAND-HAR 436
DB 511 LIHENDMMAVF--NVTAMEEKGYLOEDFEDPMNPKRAVP-----YNRDFFHAR 558
QY 437 AASIVANGLSGKITVNGQYESWPAIALSDOOIANVITYTINSFG 482
DB 559 AGNFSASBITARVQ-----ELABOEPYRRLDEILEDG 591
RESULT 17
US-10-080-233-2
; Sequence 2, Application US/10080233
; Publication No. US20020151450A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Huang
; TITLE OF INVENTION: No. US20020151450A1e1 Phenol Oxidizing Enzymes
; FILE REFERENCE: GC567
; CURRENT APPLICATION NUMBER: US/10/080,233
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Stachybotrys sp.
US-10-080-233-2

Query Match 4.4%; Score 114.5; DB 14; Length 594;
Best Local Similarity 19.0%; Pred. No. 0.092;
Matches 123; Conservative 72; Mismatches 226; Indels 225; Gaps 28;

QY 6 LIKTLICLSALMSGCSNOADKAAQPKSSTYDAAKT---ANADNASQEHGELPVI 62
DB 2 LFKSQWLAASGLSGVLGIPMDTSGHP-IEAVDPEVKTEVPADSLAAGDDMESPPY 60
QY 63 DAIVTHAPEVPP-----PV-DRD---HPAKVVMGEMTEKVMRLADGVEYQFWTF 108
DB 61 NULYNALRPIPVKQOKMIIITPVGKDIWYEIEIKPQOKRIYPTLRPATLVGD---- 116
QY 109 GGOVPGOMIRVREGDTIEVOFSNHPDSKMPHNVDFHA--ATGPGGAASFTAPGHTSTF 166
DB 117 -GMSGGPFENVPRGRETIVRFIN--NATVENSYHLHGSPSRAFPDQMAEDVTFPBGYKX 173
QY 167 SFKALQPG--LYVHCAPAVPGMHIANGMVGLIV---EPKEGLPKVDKEYYVNOGDFY 220
DB 174 YEPNYSARLLWYHDAFMKTAENNAVFQAGAVIINDAEDALGLP----- 219
QY 221 TKGKGEGLOLPFDEMEKAIREDAEVYVFNQSGALGTENA----- 260
DB 220 -SSGGE-----FDLPIL--TAKYNNADGTLRSTEGEDQMDGDIHVNGQWPPLNVO 270
QY 261 -----LKAKYGETVRLF--VGNGGPNLTSSFHVG----- 289
DB 271 PRKYRFRPLNAAVSRAAMLLYVTRTSSPNRIFPQVIASDAGLLQAPVQTSNLXLAERY 330
QY 290 --IDKVFHEGKGENHNIQTT-----LIPAG----- 314
DB 331 EIIIDFTVFAQQTLLRLNVAETNDVGDDEYARTLEVMRFVSSGTVEDNSQVPSLTDV 390
QY 315 -----GALTEFVNDVGDVYLVNDHAF-----RAFKKALGLIKVGEEN-----H 356
DB 391 PPPPKKEGADGHFPERSNHLYLNDVGFADVNEVLAKPELGTVEVLELNSGSGWSH 450
QY 357 ELYSH-----KOT-----DAVYLBEGAPQAIIDTOEAKTPAPA---N 390
DB 451 PVHILVDEKIKRTRGQVMPYESAGLKDYYWLGREGTLLTEAHYQWTAIYMHCHN 510
QY 391 LOEOIKAGKATYDSNCAACHQ-----PDGKGVNAPFPLANSDYLAND-HAR 436

DB 511 LIHENDMMAVF--NVTAMEEKGYLOEDFEDPMNPKRAVP-----YNRDFFHAR 558
QY 437 AASIVANGLSGKITVNGQYESWPAIALSDOOIANVITYTINSFG 482
DB 559 AGNFSASBITARVQ-----ELABOEPYRRLDEILEDG 591

RESULT 18
US-09-995-749A-13
; Sequence 13, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-13

Query Match 4.3%; Score 113.5; DB 10; Length 535;
Best Local Similarity 21.8%; Pred. No. 0.097;
Matches 73; Conservative 43; Mismatches 116; Indels 103; Gaps 16;

QY 165 TFSFKALQPGLYVY---HCAVAPVGMHIANGMVGLILVEPKGLPKVDKEYYVNOGDFY 220
DB 175 TTFEDEQKQIDAYIIONDSYVKKYNNIPASVAILLTN-KDTIPRV---GDLY 227
QY 221 TK-GKYGEGLOPFD-MEKAIREDAEVYVFNQSV-----GALTGN 259
DB 228 TDGQGYMEHQRTFYDTLLNLKSRVYVAGGQSMQMTSGVGNNNILTSRYGKGAMTAVD 287
QY 260 A-----LKAKYGETVRLFVGNGGPN-----LTSSFHVGIFPK 293
DB 288 TGTDETRTQIGCVVSNTPNLKLVNDKVLHMGAAHKNQYRAVLTTTDEVINTSDQ 347
QY 294 VHEGKGENHNIQTTLLPAGGAITFEKVDPVGDVYLVNDHAFRAFNKALGLIKVEGE 353
DB 348 -----GAPVA--MTDENGDLYSSNHL--VNGKEADTAVQGY 382
QY 354 ENHEIYSHKQTDVYLBEGAPQAIIDTOEAKTPAPANLOEIKAGKATYDSNCAACHQPD 413
DB 383 ANPDVSGYLV--AVWVPVQASDNDQARTAPST-----EKSGNSAYRTNAF---D 427
QY 414 GKGVENAF-----PLANSDYLANDHARAASIVAN 443
DB 428 SNVIFEAFSNFTYPTKESERANVRIQWADFFAS 462
RESULT 19
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A

CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 09/604,957
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: ERO 00201871.1
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1781
TYPE: PRF
ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match 4.3%; Score 113.5; DB 10; Length 1781;
Best Local Similarity 21.8%; Pred. No. 0.63;
Matches 73; Conservative 43; Mismatches 116; Indels 103; Gaps 16;

QY 165 TFSFKALQPLGVY---HCAVAPVGMHIANGVGLIVPEKGLPKVDKEYVVMQDFY 220
DB 1154 TTFEEDQKIDAVIQDQNSTVKKYNLXNIPASAYAILTN-KDTIPRV---YY---GDLV 1206
QY 221 TK-GKXGEOGLOPFD-MEKAIREDAYVNVGSV-----GALTGEN 259
DB 1207 TDGQYMEHQTRYDVLTLNLKSRVKYVAGQSGQWQSVGNNNLTYSVYGGKAMTATD 1266
QY 260 A-----LKAKEGETVALFVNGGSPN-----LTSFHYIGEIFDK 293
DB 1267 TGTDETRTQIGVNVSTPLKGVNDKVLHMGAAKQOYRAAVLTITDGIYNTSDQ 1326
QY 294 VHEEGKGENHNIQTLLIPAGAAITFEKVDVPGDYVLVDHAI FRAFNKALGILKEVE 353
DB 1327 -----GAPVA--MTDENGDTLSSHND--VYNGKEADTAVQGY 1361
QY 354 ENHEIYSHKOTDVLVPEGAPOALDTQEAETKAPAPALQOIQIAGKATYNSNCAQHPD 413
DB 1362 ANDVVGYL---AVWVPGVAGSDQDARTAST-----EKSGNSAVRTNAAF---D 1406
QY 414 GKGVNPAF-----PPLANSDYLANDHARAASIVAN 443
DB 1407 SNVIFAFSNFVYTPTESERAVRIRIAQNADFPAS 1441

RESULT 20
US-10-095-718-4
Sequence 4, Application US/10095718
Publication No. US20020131956A1
GENERAL INFORMATION:
APPLICANT: Walsh, Christopher
APPLICANT: Chao, Hengjun
APPLICANT: Burslein, Haim
APPLICANT: Lynch, Carmel
APPLICANT: Stepan, Tony
APPLICANT: Munson, Keith
TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
FILE REFERENCE: 35052/204375
CURRENT APPLICATION NUMBER: US/10/095,718
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/689,430
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/158,780
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1431
TYPE: PRF
ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match 4.3%; Score 113; DB 14; Length 1431;
Best Local Similarity 20.3%; Pred. No. 0.5;
Matches 82; Conservative 66; Mismatches 154; Indels 102; Gaps 22;

QY 68 HAPVPPVDRDH-----PAKVVKMEV-----EKWRLADGVEVFTFGGOVP 113
DB 408 YABSGPTPDRSHKULYLANGPQRIQKTKYKRPAAVYTBETKTRATIQESGILGPLY 467
QY 114 GQIRVREGDTIEVOFSN--HPDSKMPHNVDF---HAATPGGGAESF--TAGHTS 164
DB 468 GEV-----GDTLLIIFKNQASRPVNIYPHGINVVPRLHTGRLPKGVKHKDMPILFG--E 520
QY 165 TFSFK--ALQPG-----LVYHCAPVGMHIANGVGLIVPEKGLPK----- 207
DB 521 IFKYKWTVEEDGPTKSDRCLTRYSSPINERDLASGLIPLLCYSESVDORGNQNM 580
QY 208 VKEEYVMQDGF-YTKGKXGEOGLOPF--DMEKAIREDAYVFN--GSVALTGEN-AL 261
DB 581 SDRKAVILSVSDENRSWYLTENMORFLPNADVQPHDEPQLSNMHSINGVFPNLQ 640
QY 262 KAKVGETVALFVNGGPNLTSFHYIGELFDKVEEGKGENHNI--QTLLIPAGAAI 318
DB 641 SVCLHEVAVAYILSVQAQTD-----FLSVFSGYTFKHMVYEDTLTLPFGSETV 691
QY 319 TEFRKVDVPGDYVLVDH-AIFRAPNKALGILVEGSEHHEIYSHKOTDVLVPEGAPOAI 377
DB 692 F-MSMENPGLMWLVCNHSNDFR--NRGMTALLKVS-----CNRNIDDY-----E 733
QY 378 DTQEAETKPA-----PANTLOEIKAKATYD 403
DB 734 DTYEDIPTPLANNVNIKPRFSQNSRHPSTYKQOLKMKREDP 777

RESULT 21
US-10-174-693-290
Sequence 290, Application US/10174693
Publication No. US20030131373A1
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.10035
CURRENT APPLICATION NUMBER: US/10/174,693
PRIOR FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
PRIOR APPLICATION NUMBER: US 09/615,192
PRIOR FILING DATE: 2000-07-12
NUMBER OF SEQ ID NOS: 407
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 290
LENGTH: 440
TYPE: PRF
ORGANISM: Eucalyptus grandis
US-10-174-693-290

Query Match 4.2%; Score 110.5; DB 16; Length 440;
Best Local Similarity 19.7%; Pred. No. 0.14;
Matches 86; Conservative 51; Mismatches 136; Indels 163; Gaps 20;

QY 111 QYRGKIRVREGDTIEVQSFNSHPDKMPPHNVDF-----ANTGCGGAFAFTAPGHTS 164
DB 75 QYGPPLVAREGDNILVKVNH---VAANVTIHHGVROLRTGMADGAVYTQCIQTN 130
QY 165 ---TFSFKAL--OPGLVYVHCAVAPVGMHIANGVGLIVPEK--EGLP--KYDKEYVVMQ 216
DB 131 QSYTNFTLTGQRGLTMAHVS---WRRSIHGPITLIPKRNESYPREKSKVPIIF 186
QY 217 GDFYTKGKXGEOGLOPFDMEKAIRED-----AEYVFNQSGVGLTGENA-----LK 262
DB 187 GEMFN-----VPEAVIAQALQSGGPNVSDATYINGLPGPLVNCSSKDTFKLK 235

QY 263 AKVGETRLPVNGGNLTSSPHV-----IGEIFDK-----VFEGKGENHNIO 307
 DB 236 VPRGKYLLRLINAAINDELFFSIANNAVVEVDAYTKPFSAGCLHLPQGTMMVLK 295
 QY 308 TLPIAGGAATIEFKYDVPDGYLVDAIFRAF-----NKALGILKVEGEN----- 355
 DB 296 T-----KTDFPNSTFLM--AAMPYFTGKTGFDNSTVAGILEYEHKSSNYP 340
 QY 356 -HEIYSHKOT-----DAVYLPEGAPOAID-----TOEAPKTAPA 389
 DB 341 LKKLPQYKTLPPMNSTGVAKTQGLRSLASAKFPAANPQKDKRFFFTVGLGTSPCR 400
 QY 390 NLQEOIKAKATYDSCAACHPDGKGVNAPFPPLANSDYLNADHARAASIVANGLSGKI 449
 DB 401 N-----TTCQGPNGT-----KF 412
 QY 450 TVNGNOYESVMPAL 465
 DB 413 AASVNNISFVLPSVAL 428

RESULT 22

US-09-118-276-2
 ; Sequence 2, Application US/09118276
 ; Patent No. US20010011381A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BABYCHUK, ELENA;
 ; APPLICANT: KUSHNIR, SERGEI;
 ; APPLICANT: DE BLOCK, MARC;
 ; APPLICANT: INZE, DIRK
 ; TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SIXBEY, FRIEDMAN, LEBDOM, & FERGUSON
 ; STREET: 8180 GREENSBORO DRIVE, SUITE 800
 ; CITY: MCLEAN,
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3-1/2" DISKETTE
 ; COMPUTER: IBM-COMPATIBLE
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/118,276
 ; FILING DATE: 17-JUL-1998
 ; PRIOR APPLICATION DATA: NONE
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
 ; REGISTRATION NUMBER: 31,196; 43,077
 ; REFERENCE/DOCKET NUMBER: 6201-0003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 790-9110
 ; TELEFAX: (703) 883-0370
 ; INFORMATION FOR SEQ. ID NO. 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 969 RESIDUES
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: SINGLE
 ; TOPOLOGY: LINEAR
 ; US-09-118-276-2

Query Match 4.2%; Score 109; DB 9; Length 969;
 Best Local Similarity 18.8%; Pred. No. 0.64;
 Matches 119; Conservative 72; Mismatches 188; Indels 254; Gaps 25;

QY 24 SNQADAAQPKSTVVA-AAKTANADNMSQ---EHQGLPVDALVTARPEVPPVDND 79
 DB 175 NNEQNGSKRKXSENDIDSKARLDESTSEGTVRNKGOL-----VDPKGSNT 222

QY 80 HPAKVTVKA-ETVEKMYRLADGEVEYQFTFGQVPGQMIRVBGDTIEVOFSNHPDSKMP 138
 DB 223 SSADIQTLKKEOSDTLMKLDKIK-----THVSAAEIRDMLEANGQDTSGB 268
 QY 139 --HNVDENHATGPGGAASFTAPGHTSTFSFKALOPGLVYVHCANAPGMHANGMYGL 196
 DB 269 ERHLID-RCADG-----MIFGALGP-----CPVCANGMYTNNQYQC 304
 QY 197 -----ILVEPREGLPKYDKEYVYVQDGFYTKGKGEOGLQFPDMEKA- 238
 DB 305 SGVSEMSKCTYSATPEPVAVKKKMQOIPHGTNDYLMKMKSKQKVKPERVLPPMSPEKGS 364
 QY 239 ----- 238
 DB 365 SKATORSTLSKGLDKLRFVVGQSKKANEMIEKTLKAGANFYARVVKDIDCLACGE 424
 QY 239 -IREDAEY-----VFNQSVGALTO-----ENAKAKYGETRLPVNG 276
 DB 425 LNNENAEVRKARLKIPIVREGYIGCVKKNRMLPDLKYLENALESKSGSTVTVKVG- 483
 QY 277 GPNLTSSFVHIGEIFDKVH-FEGKGKGNHIOITLIPAGGAATIEFKVDVPGDYLV- 332
 DB 484 ----RSAYHSSGLQDTNAILLEDGS-----IYNATLMSDLALGVNSYVQIIE 530
 QY 333 --DHAIFRAFNK-GALGILKVEGENHEIYSHKQTDV----- 367
 DB 531 ODDGSECVYFRKMGVGRGSEKIGQKLEEM---SKTEAIEKFLRPLEKTGNSWEAMECKT 587
 QY 368 -----YLPEGAPOAIDTOEAPKTAPANIQ-----EOTRAGKATYDS 404
 DB 588 NRRKOPGRFPYPLVDYGVKAPKARKDISEMKSLAPQLLELMKMLFNVTYTRAAWMEFE 647
 QY 405 NCAACHQPDGK---GVNAPFPPLANSDYLNADHARAASIVANGLSGKITVNGNOYESVA 460
 DB 648 NMS--EMPLGKSKENIEKGFALTEIONLMDTDQALAVESL---IVAASNREFTLI 702
 QY 461 PAI-----ALSDQJANYI 474
 DB 703 PSIHPIHIEDDMLMKAKMLELQDIEIASKI 735

RESULT 23

US-09-118-276-11
 ; Sequence 11, Application US/09118276
 ; Patent No. US20010011381A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BABYCHUK, ELENA;
 ; APPLICANT: KUSHNIR, SERGEI;
 ; APPLICANT: DE BLOCK, MARC;
 ; APPLICANT: INZE, DIRK
 ; TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SIXBEY, FRIEDMAN, LEBDOM, & FERGUSON
 ; STREET: 8180 GREENSBORO DRIVE, SUITE 800
 ; CITY: MCLEAN,
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3-1/2" DISKETTE
 ; COMPUTER: IBM-COMPATIBLE
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/118,276
 ; FILING DATE: 17-JUL-1998
 ; PRIOR APPLICATION DATA: NONE
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
 ; REGISTRATION NUMBER: 31,196; 43,077
 ; REFERENCE/DOCKET NUMBER: 6201-0003

```

1 TELECOMMUNICATION INFORMATION
2 TELEPHONE: (703) 790-9110
3 TELEFAX: (703) 883-0370
4
5 INFORMATION FOR SEQ ID NO: 11:
6
7     SEQUENCE CHARACTERISTICS
8     LENGTH: 980 RESIDUES
9     TYPE: AMINO ACID
10    STANDARDS: SINGLE
11    TOPOLOGY: LINEAR
12
13 US-09-118-276-11

```

Query Match	4.2%	Score 109;	DB 9;	Length 980;
Best Local Similarity	18.8%	Pred. No. 0.65		
Matches 119;	Conservative	72;	Mismatches 188;	Indels 254;
				Gaps 25;

```

RESULT 24
US-09-874-069--4
; Sequence 4, Application US/09874069
; Patent No. US20020064826A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Cytokine Receptor-like Polynucleotides, Polypeptides, and Antibod
; FILE REFERENCE: PTO21P1
; CURRENT APPLICATION NUMBER: US/09/874,069
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32525
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/168,621

```

```

; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 988
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-874-066-4

```

Query Match	4.1%	SCORE 108.5	DB 9	Length 988	
Best Local Similarity	19.0%	Pred. No. 0.74			
Matches 103; Conservative	75;	Mismatches 182;	Indels 181;	Gaps 25	

RESULT 25
 US-09-769-787-2
 : Sequence 2, Application US/09769787
 : Publication No. US20030091577A1
 : GENERAL INFORMATION:
 : APPLICANT: Microbial Technics, Limited
 : APPLICANT: Christophe, Christophe FG
 : APPLICANT: Hansbro, Philip M
 : TITLE OF INVENTION: Proteins
 : FILE REFERENCE: PMC/P212950
 : CURRENT APPLICATION NUMBER: US/09/769,787
 : CURRENT FILING DATE: 2001-01-26
 : PRIOR APPLICATION NUMBER: GB 9815337.1
 : PRIOR FILING DATE: 1998-03-27
 : PRIOR APPLICATION NUMBER: US 60/125164
 : PRIOR FILING DATE: 1999-03-19
 : NUMBER OF SEQ ID NOS: 388
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 2

LENGTH: 2233
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-787-2

Query Match 4.1%; Score 108.5; DB 11; Length 2233;
Best Local Similarity 20.3%; Pred. No. 2.6;
Matches 110; Conservative 63; Mismatches 179; Indels 189; Gaps 30;

QY 58 ELPVDAIVTHAPEVP-----PVDKDHPAKVVK----- 87
DB 994 EYVKQTIIGAEPEPTTVPFYSDGSRAREPVTWSSVDVSRGIVTVKGMADGREVEAR 1053
QY 88 -----METEYKVR--LADG--VEYOF--WTFGGQ-----VPG 114
DB 1054 VEVIKSELVVKRIAPNTDLSVDKSVYLDIGSVVEYVDKMEIAEDKAKLAIFG 1113
QY 115 QMIRV---EGD---TIEVQFSNHPDSKMPHNVDFHATGPGGGAELSFTA--PGHIST 165
DB 1114 SRIQATGYLEGPPIHATLVVEEGNPAPAVP-----TVTVGGEAVTGLTSQKMQYRT 1166
QY 166 FSKALQPL--YVHCAPVGMHIANGMVGLLVEPKG-----LPKV-- 208
DB 1167 LAYGAKLPEVITASAGNAATVTLQASANGMASIFIQKDGGLQTVYALQLEAPKIAH 1226
QY 209 -----DKRYVMQD--FYTKGXGE-----QGLQPFDM 235
DB 1227 LSLQEKADSLKEDQTVKLSVRAHYQDGTQAVLPADKVTFTSGE--GEVAIRKMLELHK 1285
QY 236 EKAIREDAEYVFNFSVGVLTGENALKAKVGETVRLFVNGGNPLTSSPHVIGEIFDKVH 295
DB 1286 PGAVTLNMEYEGAKQV--ELTIQANTEKKIAQSIR-----PVNVVTDLHQEPSLPATVT 1338
QY 296 FEGGKG--ENHIQITLLPAGGAATTEPKVDVPGDVLVDHAIIFRAFNKGAIGILKVEE 353
DB 1339 VEYDGFPTKHTVTQALP-----KEKLD-----SYQFFB--VLG--KVBGI 1376
QY 354 ENHEIYSHKQTDVAVLPFGAPQALDIOEAPKTPAPANLQEQIKAGKATYDSCAACHOPD 413
DB 1377 D-LEARAKVSVGIVSEVSTTPIAEAPQLP-----ESVR-----TIDSN--GHVS 1422
QY 414 GKGVNAEPPLANSOYLADHAPASIVANGLSKITVNGNOYESVMA--IALSDQOLAN 472
DB 1423 AKVADAIRP-----EQYAKEGVFTVNG-----RLEGTQLTKLHVRVSAQTEOGAN 1469
QY 473 V 473
DB 1470 I 1470

RESULT 26
US-09-765-272-84
Sequence 84, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/361,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 722 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-765-272-84

Query Match 4.1%; Score 108; DB 9; Length 722;
Best Local Similarity 19.8%; Pred. No. 0.5; Indels 180; Gaps 27;
Matches 104; Conservative 65; Mismatches 176;

QY 58 ELPVDAIVTHAPEVP-----PVDKDHPAKVVK----- 87
DB 263 EYVKQTIIGAEPEPTTVPFYSDGSRAREPVTWSSVDVSRGIVTVKGMADGREVEAR 322
QY 88 -----METEYKVR--LADG--VEYOF--WTFGGQ-----VPG 114
DB 323 VEVIKSELVVKRIAPNTDLSVDKSVYLDIGSVVEYVDKMEIAEDKAKLAIFG 382
QY 115 QMIRV---EGD---TIEVQFSNHPDSKMPHNVDFHATGPGGGAELSFTA--PGHIST 165
DB 383 SRIQATGYLEGPPIHATLVVEEGNPAPAVP-----TVTVGGEAVTGLTSQKMQYRT 382
QY 166 FSKALQPL--YVHCAPVGMHIANGMVGLLVEPKG-----LPKV-- 223
DB 436 LAYGAKLPEVITASAGNAATVTLQASANGMASIFIQKDG----- 478
QY 224 KYGEGLOPFMEKAIREDAEYVFNFSVGVLTGENALKAKVGETVRLFVNGGNPLTSS 283
DB 479 -----LQTVAIQ--FLEAPKIAHL-----SLQVEKADSLKEDQTVKLSV----- 516
QY 284 FHVIGEIFDKVHFEGGKENNIIQITLLPAGGAATTEPKVDVPGDVLVDHAIIFRAFNK 343
DB 517 -----RAHYQD-----TOAVLPADKVT--FSTSGE--VAIRKMLELHK 556
QY 344 ALGI-LKVEGEENH-EIYSHKQTDVAVLPFGAPQAL--DIOEAPKTPAPANLQEQIKAG 399
DB 557 AVTLNMEYEGAKQVDELTIQANTEKKIAQSIRPVNVVTDLHQEPSLPATVVE----- 609
QY 400 ATYDSCAACHQPDGKVP-----NAPPLANSOYLADHAPASIVANGLSKITVY 452
DB 610 --YDGFPTKHTVTQALPKELDLSYQTFEVLGKVEGIDLE-ARA-----KVSVE 656
QY 453 G---NQESVNPATIALSDQOLANVTITLNSFGNGGQSLADVA 494
DB 657 GIVSVSEVSVTPIAEAPQLPESVTTYSN-----GHVSSAKVA 695

RESULT 27
US-10-238-075-877
Sequence 877, Application US/10238075
Publication No. US20030148324A1
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
FILE REFERENCE: B.LANDINE
CURRENT APPLICATION NUMBER: US/10/238,075
FILING DATE: 2002-09-10

PRIOR APPLICATION NUMBER: 0003145
 PRIOR FILING DATE: 2000-03-10
 NUMBER OF SEQ ID NOS: 1576
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 877
 LENGTH: 1684
 TYPE: PRT
 ORGANISM: Escherichia coli
 US-10-238-075-877

Query Match 4.1%; Score 108; DB 12; Length 1684;
 Best Local Similarity 20.3%; Pred. No. 1.9; Indels 224; Gaps 33;
 Matches 127; Conservative 76; Mismatches 199;

5 TLTKTLTLCALSLMT-----SGCSNO--ADKAQPKSSTV 38
 63 TLNSALTYAGNNLKFTBRHNOHGDILADNSLMQKSSGGRANSEIIRNSNIETMG 122
 39 DAAAKTA---NADNAASQEHQGLPVIDAVTHAPEVPPVDRDHPAKVVMKETEYKVM 95
 123 DITMNTAHLNSWDALISASH-----EVIKSSRGVISPVENNMWGVV----- 167
 96 RLADGVEY--QFTWPGQVPGQ--MIRVEDTIEVQSNHPSDKMHNVDFAATGPGGG 152
 168 --HDGEYLAIVWGKATVPDEYRIRGTDTVTATGHA--ARISGADMHIRARLDN 224
 153 ABAFPAQHTSTPSEFKAL-----OPGL-----YVHCVAVAVGMHIANGMVGLIVEPK 203
 225 -EASFILAGSMTLSGDTLNNQMGOSTCKEYVWLA-----SD 263
 204 GLPKVDKEYVNVQGDFTYKGY-----GEOGLOPFMEKAIREDAEVYVNGSV 252
 264 SLPKA-----WTFEPWYVVRQVSPDATEAGTSPVQYRAVISAAGV--SASF 311
 253 GALTGNALKAKGVEVRLFVNGG---PNLTS-----SFVIGIEIPKVFEGKG- 301
 312 ATDTGNTTVPRAQ-----GSGNTITVPSLNSLTPPTVSQSGSEALLN--ESGTGI 361
 302 ----EHNHIOITTL--IPAG---GAATFEP-----KVDPVGDVVL----- 331
 362 TGPVWMDALPDTLKDPGALSILGASVSYPLPSGNNGVFPSTDDSPILITVNRKLG 421
 332 ---VDHAIFRAFNKGAIGILKYGEENHEIYSHKQTDVYLPE---GAPQALDTGEAPK 384
 422 LGKVDSLF---AGLYDLRMQGE---APRETDPVYIDKQFLGSSYIIDLRLGPK 472
 385 TPAPALQGIKAKGKATYDSNCAACHQPDCKGVPNAPFLANSDYLNAD----- 433
 473 E-----KDYRFLGDAAFDT-----RYSNVILNQTSRYINGTGSDDLTKMKYLM 516
 434 HARAASIVANGLSGKITVNGNQ-----YESV-----MPAIALSDQOI----- 470
 517 DSAALAAQKALGLTFGSLTAGOVAQLTRSLMWESVITINGQYMWVKLYISPDITLHNG 576
 471 ---ANVTYTLNLSFGNKGGLSADD 492
 577 SVISGNNVQLAGNITNSGGSINAON 602

RESULT 28
 US-09-738-363-6
 Sequence 6, Application US/09738363
 Patent No. US20010010932A1
 GENERAL INFORMATION:

APPLICANT: Schmeff, Harry E.
 Schmeff, George E.
 Payne, Jewel M.
 Narva, Kenneth E.
 Foncerada, Luis
 TITLE OF INVENTION: Nematocidal Proteins
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESS: Jay M. Sanders

STREET: 2421 N.W. 41st Street, Suite A-1
 CITY: Gainesville
 STATE: FL
 COUNTRY: USA
 ZIP: 32606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA: US/09/738,363
 APPLICATION NUMBER: US/09/738,363
 FILING DATE: 15-Dec-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/076,137
 FILING DATE: 12-MAY-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Sanders, Jay
 REGISTRATION NUMBER: 39,355
 REFERENCE/DOCKET NUMBER: MA-20CCCD3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 352-375-8100
 TELEFAX: 352-372-5800
 INFORMATION FOR SEQ ID NO: 6 (PS33F2):
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1257 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Bacillus thuringiensis
 INDIVIDUAL ISOLATE: PS33F2
 CLONE: E. coli NM522 (pMYC 2316) B-18785
 FEATURE:
 NAME/KEY: Protein
 LOCATION: 1..1257
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-09-738-363-6
 Query Match 4.1%; Score 107.5; DB 9; Length 1257;
 Best Local Similarity 19.2%; Pred. No. 1.3;
 Matches 118; Conservative 69; Mismatches 200; Indels 227; Gaps 32;

25 NOADKAAQPKSSTVDAAKTANADNAASQEHQGLPVIDAVTHAPEVPPVDRDHPATV 84
 245 NGLDKF---KSLDVSYNKAKAVIKGMB-----MVLDLVA-----LWTFDDPDHYQKE 290
 85 VVKMETVEKMLADGVEYQ-----FWTFGGQVPGQMLRVNEDTIEVQFSNHPDS--- 135
 291 V---EIEFRTTISPI-YQVPVKMONTSSSIVPSDLFHY-QGDVLKLEFSTRIDNDGL 344
 136 -----KMP-----HNVDFAATGPGGAAASFTAP----- 160
 345 AKLFTGIRNTPFKSPRTHETVHDFSINTOSSGNISRGSSNPIPIDLNPIITCTIRNSF 404
 161 -----GHTSTFSPKALQPGLYVYHCAVAPVG-----MHIANGY----- 194
 405 YKALGSSVILVNFK---DGTGYAFAPQAPTGAMDHSEFSDGAPRGHKLNTYTSPPGT 461
 195 -----GLIVEPKGGLPVVDKEYVYMQG--DFTYTGKKGEGGLQPFME 236
 462 LRDFINVTLLSTPFINELSTEKINGFP-AEKGYIKNQGIKMYKRPY-INGAOPVNLB 519
 237 KA-----IREDAEYVY-----FNGSVGLTGENA-----LKAIVGET 268
 520 NQOTLIFEPHASKIAQYIRIRYASTQGTGKGFRLDNDGLQTLNPTSANGVYTNIGEN 579
 269 VRLFVNGGPNLTSSFHYI-----GEIPDKVHFEGGKGNNHNIQTTLLPAGGAATTE 320

Db 580 YDLYT-IGSYTTEGNTLQIOHNDKNGVLDRIEF-----VPSDSLQ--- 621
Qy 321 FKVDVGDVYVLDVHAFRAFNKCALGLKVEGENEIIYSHKQTDVYLPEGAPOAIDTQ 380
Db 622 ---DSQDDEPPEHETIIFDKSSPTI---MSNGSHSHILESTYSQSS----- 667
Qy 381 EAPKTPAPANLOEQIKAGATYDSNCACHQDPDGKVPNAPPL-ANSDYLNADHARAAS 439
Db 668 -----YPHNLLINL-----PHPTD-----PRNHTIHVNNQDMVNDYCKDS- 703
Qy 440 IVANGI-----SGKIT-----VNGNOYESVMPALALSDQOIANVITYTLNS 480
Db 704 -VADGINTFKKITATIPSDAMYSGLTSMHLEFNDNFKITTPKRELSN-ELENTITQVNAL 761
Qy 481 FGNKGQSLADVA 494
Db 762 PASSADOTLASNV 775

RESULT 29

US-10-043-487-383
; Sequence 383, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: PIERRE, LEOGRAIN
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: protein-protein interactions between Shigella Flexneri polypeptide
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: Patent version 3.1
; SEQ ID NO 383
; LENGTH: 2609
; TYPE: PRT
; ORGANISM: Shigella Flexneri
US-10-043-487-383

Query Match 4.1%; Score 107.5; DB 15; Length 2609;
Best Local Similarity 20.8%; Pred. No. 4.1;
Matches 115; Conservative 71; Mismatches 217; Indels 149; Gaps 30;

Qy 4 PRLITTLICALSALMLSCGNOADKAAQPKSYDAAKTANADNAAOEHOGELPID 63
Db 2113 PSVANVGSCHDLS-LKPEISIQ-DWTAQVTS---PSGKTHAEIIVEGENHTYCIKRV 2166
Qy 64 AIV-----THAPEVP-----PRVDRDHPAKVVVVKMETVEKVMILADGV-EYGF 105
Db 2167 AEMGHHTVSVKKGQVHPGSPPOFTVGLBEGGANHVRAAGPGLL---PAEAGVPAPFETI 2223
Qy 106 WIFGQVPGQMITRVEBDTIEVOFSNHPDS-----KMPHNVDFHAATGCGGAEASF 157
Db 2224 WTRKAGAGLALAVGSPKAEISFEDRKQSGCVAVVQEPDGYEVSVFNEHPIRDSF 2283
Qy 158 TAP-----GHSTFSFKALQ-PGLYVTHCAVAPVGNHIANGYGLILVEPKGELPKVDKE 211
Db 2284 VVPVAPSGDARRLTVSSLSQESGLKVNQAPASPAVSLNGAKAIDAKVHSPSGL---SE 2339
Qy 212 YVVMGQDFYTKGKYGGQGLQPFDMKAIREDAEYVV---FNGSVGLTGENMLKAKVGET 268
Db 2340 CVYTEID---QDKAVRFIP-----RENGVYLLIDVFNGL--HIPG-SPKIRVGP 2385
Qy 269 VALFVNGG-PVLTSSFHVIGEIFDKVHFEGGGEN---HNIQTLLIPAGGAIT----- 319
Db 2386 ---GHGDPGLVSAV-----GAGLEGVTVGNPAFVNTSAGAGALSVITDGS 2432
Qy 320 EKVVD-----VQGDYVL-----VDHAFRAFNKCALGLKVEGE-- 353
Db 2433 KVMDCQECBPGRVTVYTPMAPGSYLISIKYGGPYHIGSGPPKA-----KVTGPR 2483

Qy 354 -ENHEIYSHKQTDVYVLEPAGAPQIDTOEAPKTPAPANLOEOIKAG-----KATY 402
Db 2484 VSNHSLH---ELSSVFWDSLTATCAPOHGAFCGPBQADSKVAVANGLSKAYVQKSSF 2540
Qy 403 DSNCA-----ACHQD---DGKVPNAPPLANSDYLNADHARAASIYANGLSKI 449
Db 2541 TVDCSKAGNNMLVGVHGPRTPCEIILVGVHSRLYSVYLLKDXGEYTLVVK---WGHE 2597
Qy 450 TVNGNOYESVMP 461
Db 2598 HTPGSPYRVVP 2609

RESULT 30

US-10-156-761-10292
; Sequence 10292, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10292
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10292

Query Match 4.1%; Score 106.5; DB 15; Length 482;
Best Local Similarity 22.7%; Pred. No. 0.37;
Matches 85; Conservative 46; Mismatches 143; Indels 101; Gaps 20;

Qy 114 GQMITRVEBDTIEVOFSNHPDSKMPHNVDFHAATGP---GGAEASFAPRPHGT--STSEF 168
Db 88 GPTMKWTSQDVTYLLNITNSLADT--TVHFQAHIPKXQDGPONAFNA-GETWSPTFTV 144
Qy 169 KALQGLVYVHCAVAPVGNHIANGYGLIIVE---PKGELPK---VDKEYVVMQ----- 216
Db 145 KDEAKTLWHHPALSTTBEOYTRGLAGMIYEDSDSAAALSTYGTDDIPILLOCLAAD 204
Qy 217 --GDF-YTKGKYGGQGLQPFDMKAIREDAEYVNGSVGALTGENALKAKVGETVRLFV 273
Db 205 SSGDVKYVLDYLSGGLSVPLCNGTNDVDTTLAFT-ATKSTRRLNLAASPSDITVQR 263
Qy 274 GNGGNLTSFHVIGEIFDKVHFEGGKBNHNIQTLLIPAGGAITTEKRVNDVPGDYVLD 333
Db 264 GGG--TLTQI-----ATDQYLTAEAEVESIRLVAGARA--EFMDL----- 302
Qy 334 HAIFRAFNKCALGLKVEGENEIIYSHKQTDVYVLEPAGAPQIDTOEAPKTPAPANLQ 393
Db 303 -----SDAVTL-----QAVVT-----TG 315
Qy 394 QIKAGATYDSNCACHQDPDGKVPNAPPLAN--SDYLNADHARAASIYANGLSG-KIT 450
Db 316 WVRGSGTYDFLTV---PDASDTPDALPSSLNITTRVDTTDF--AARTITLGGSGASML 370
Qy 451 VNGNOYESVMPALAL 465
Db 371 INGSA-GTTMAAMM 384

US-10-156-761-14797

Query Match 4.0%; Score 104.5; DB 15; Length 625;
Best Local Similarity 20.0%; Pred. No. 0.85;
Matches 111; Conservative 55; Mismatches 191; Indels 197; Gaps 28;

QY 19 MLSGSSNOADKRAQPKSSTVDAAKTANADNAAQHEGELFVIDAIVTHAEVPPVDR 78
DB 22 LIGGATSPA--AAAPVTPVVVDVDDYCADPTGRTD---STPVAALHAKSV----- 70
QY 79 DHPRAVVMVMEVEKVMRLADVEYQFWTFGGQVPCQMIRVREGTIEVFQFN----- 131
DB 71 DRPVIVFSKGT-----YQLYP-----BRAETRELMSNTVYGADOR 106
QY 132 HPDSKPMHV--DFHAATGPGGAE-----ASTAPGHT-STFSFKALQPGLYVH 179
DB 107 YRDKKIGLLVEDMHVYTDGSGAKLVHGLQTAFAISIRSTDTVFQNFSDYAP--EVID 164
QY 180 CAVAPVGMHIANGMVGLILVEPKGLPKVDKEYVMQGF-----YTKGYGSGLOP 232
DB 165 ATVATTG--VTDGHAIVRLKIPAGSPRYNGTHITWLGETSPATGQPYWGVDLQYTG 222
QY 223 FDMKALIEDBAEYVFNQSVGALITGENLAKAVGEIVLFFVNGG----- 277
DB 223 HDPEAQRTWRGDNPLFN-DVAAVTDLGGRRIRIDYTTARPADAGLVYQMLIERTEPGA 281
QY 278 -----PNLT-----SSFHVGEL-----FDKVHFEKGKGENHNIQTLLIPAGGAAI 318
DB 222 FIMSKVNTMRSMAYILOSFGVVGQFSNISIKNFAPD-----PNSGRST 329
QY 319 TEFKVDVQGVYLVDAIFRAFNKGLILK--VEGENHEIYSHKQTDAYVLPBGAPOA 376
DB 330 ASF-----ADFQMSGV-----KGVKSITRSLFDQPHDPINIH---GTYL----- 367
QY 377 IDTQAPKTPRANIOEQIKAGKATYDSNCAACHQPDGCVNAP-----P 423
DB 368 -----EVGKPKGPSTL-----TLAYKHPQTAGFPQAFGDEVEPATKRTMTP 409
QY 424 LANSPLYNADHRAASIVANGLSK-----ITVNGQYESV--MP 461
DB 410 LADA-----HAQVYAV--DGPSGMDHTKPLTTMTVTTPRPVAGVETGIVENTATP 461
QY 462 AIALSDQOIANYIT 475
DB 462 SVVISGNVFRNVPT 475

RESULT 34

US-10-137-870-234
; Sequence 234, Application US/10137870
; Publication No. US20030138883A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvarioff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C155
; CURRENT APPLICATION NUMBER: US/10/137,870

; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-870-234

Query Match 4.0%; Score 104; DB 12; Length 1160;
Best Local Similarity 19.7%; Pred. No. 2.5; Indels 108; Gaps 20;
Matches 69; Conservative 58; Mismatches 116;

QY 102 EYQWTFPGQV-----GQIRVREGDTIEVQSNHPSKMPHNVDFHAATG 148
DB 791 EYTDGTF--RIPRPTGPEHGLIGLPIKGVGILVYFNK--NASRPSVH-- 842
QY 149 PGGGAASFTY-----PGHTSTFSFKALQ--PG-----LVYHCAVAPVGMHIAN 192
DB 843 ---GVLESTTWPLAEBGCVVYQMNIPERSGPGPNDACVSWIYSAVDPI--KDMVSG 898
QY 193 MYGLI-----LVEPKGLPKVDKEYVM-----QSDPYTKGKYGEGLOPFDMKAR 240
DB 899 LVGPLAIQCKGLEPHGRSDMDREFALLFLIEDENKSMYLEENVATHGSO--DQGSINL 956
QY 241 EDAEYVFNQSVGALTGE-----NALKAKVGETVLFVNGGPNL---TSFHVIGELPD 292
DB 957 QDETLEEN--KHAHNGGLXANLRGLTMYQSERVAMWMLAMQDVLHITHHASFLL- 1014
QY 293 KVHFEKGKGENHNIQTLLIPAGGAAITEFKVDVQGVYLV--VDH-----AIFRAENK 342
DB 1015 -----RNGENVRADVDLFPCTFEVEMVASNPGTWMHCHVDHVAHMETLFTVFSR 1068
QY 343 GALGILKVEGENHEIYSHKQTDAYVLPBGAPOAIDTQAPKTPRANIOE 393
DB 1069 -----TEHL-----SPLTVITKETEKP--PRDIE 1092

RESULT 35

US-10-140-018-234
; Sequence 234, Application US/10140018
; Publication No. US20030138885A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvarioff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C158
; CURRENT APPLICATION NUMBER: US/10/140,018
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-018-234

Query Match	4.0%	Score 104	DB 12	Length 1160
Best Local Similarity	19.7%	Pred. No. 2.5	Indels 108	Gaps 20
Matches	69	Conservative	58	Mismatches 116
Qy	102	EYQWTFGGQVDP-----	GOMIRVREGDITIEVQSNHSDSKMHPVDFHAAATG	148
Db	791	EYTDGTF--RIPRPRGPENHIGLIPRIKGEVGDILTVFKN--NABRPYSVNAH-----		842
Qy	149	PGGGAEASTTA-----	PGHSTSEFFKALQ--PG-----	LYYHCANAVGCHNANG 192
Db	843	---GLESSTVWPLAEPGEVVTYQMTIPERSGPGPNDSACVSMIYSAVDPI--KDMVSG 898		
Qy	193	MYGLI-----	LVEPRKGLPKYDKKEYYW-----	QGDPLYTKGKGEGGLQPFDMKAIK 240
Db	899	LVGPRALICOKGILHEHGGKSDHDFRALLFLIFDENKSYLEENVAATGSGO--DVGSLN 956		
Qy	241	EDAEVYVFNQSVGALTGE----	NALKAKVGETVRLFYVNGGGRNL--	TSSHFVIGLEIFD 292
Db	957	QDETLESN--KMHAINGKLYANLRGLTWQGERVAMMYMLMAGQVDLTHIHNAESFLY- 1014		
Qy	293	KVHEFGGKGEENNIOTTLIPAGGAALTEPKVDVPGDYVL----	VDH-----	ALPRAFNK 342
Db	1015	-----RNGENYRADVVDLPFGTFEVEVEMVAVNSPQTMIMHCHVTDHVAGMETLETFVSR 1068		
Qy	343	GALGILTKYGEENHEIYSHKOTDAUYLRBVGARQALDQEARTPARALQ 393		
Db	1069	-----TEH--	---SPLVITKETEKEVP--PRIIE 1092	

```

RESULT 36
US-10-140-021-234
; Sequence 234, Application US/10140021
; Publication No. US2003013886A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Deonoyers, Luc
; APPLICANT: Flivaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zhenli
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C167
; CURRENT APPLICATION NUMBER: US/10/140,021
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-021-234

```

Query Match	4.0%	Score 104	DB 12	Length 1160
Best Local Similarity	19.7%	Pred. No. 2.5		
Matches 69	Conservative 58	Mismatches 116	Indels 108	Gaps 20
OY	102	EQQTFFGQGV-----GQIRAREGPTIEVQSNHDSCKPHNVDPHATG	148	
		:::		
DB	791	EYTDGTF--RIIPRPTGPEHILGILGPIIKGEVIGILTVFKN--NAPRPVVAH----	842	
OY	149	PGGAAGASFTA-----PGHTSTFSFKALQ-----LVYHCAVAPVGMHIANG	192	

```

Db      843  ---GVLESTTWPLAEBGEVTVTQWNIPEBSGGPPNDSACVMTIYSADVPI-KDMYSG  898
      193  MYGL-----LVPRKGLPKVDNEYVM-----QGDFTYKKGKKEGGLPDMEKAIR  240
Db      899  LVGPALICQCKITLPHGGSDMDPEFALLPILPEBKNSMYLEENVATHGSC--DGSITNL  956
Qy      241  EDAEYVFNQSVGALTGE---NALKAKVGETVRLPVGNGGPNL---TSSPHVIGETIFD  292
Db      957  QDEFFLESN--EMAHINGKLVTANLGLTMYQGERPAMVWYLLAMGQDVDTHTTFHESFLY-  1014
Qy      293  KVFHEGGEGSEHNHQTLLIPAGGAITEFKVDVGDVYL---VDH-----AIPAEFNK  342
Db      1015  -----RNGENRYRADVVDLPFGTEFEVEMVANSNGTMTLMCHCHVDHVAAGMETLFTVSR  1066
Qy      343  GALTGLKVEGSEHNHETSHKQTDVAVLPBEGAPQALIDQEAKTAPANLQE  393
Db      1069  -----TEHL-----SPLTVIKETEKVP-PRTEIE  1092

```

```

RESULT 37
US-10-140-274-234
; Sequence 234, Application US/10140274
; Publication No. US20030143674A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C161
; CURRENT APPLICATION NUMBER: US/10/140,274
; PRIORITY FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-274-234

```

Query Match	4.0%;	Score 104;	DB 12;	Length 1160;
Best Local Similarity	19.7%;	Pred. No. 2.5;		
Matches 69;	Conservative 58;	Mismatches 116;	Indels 108;	Gaps 20;
QY	102	EYQWTEGGQV-----	GMIRVREDTIEVGSNRPDKMNNDFHATG	148
		:::	:::	
Db	791	EYDGTG--PIPRPTGPEEHLGLGLKEVDILLVFKN--NMRPSTVANH---		842
QY	149	PGGGAESAFTA-----	PGHTSTFSFALQ---PG-----	192
		:::	:::	
Db	843	---GVLESTTWWPLAAPGEVATYQWNIPEKSGGPDNDSACVSIYTSAYDPI--KDMYG		898
QY	193	MYGLI-----	LYVEKKEGLPKVDVEYVM-----	240
		:::	:::	
Db	899	LVGEPLALCQKGLIEPHGGRSDMDREFPLPIIPENKSWYLEEVAVHGSQ--DPGSLNL		956
QY	241	EDDEYVFNQSVGALTGE-----	NALKAVGETYRLTFVNGGPNL---TTSFHYIGELFD	292
		:::	:::	

```
Db 957 QDEFLEEN-KHAIINGKLYANRLGLTMYQSERVAMWMLAMGQDVDLHTIHFHASEFLY- 1014
Qy 293 KVHFEKGGENNIOTTLIPAGGAAITEFKVDPGDYVL-----VDH-----AIFRAFNK 342
Db 1015 -----RNGENYRADVDLFPGTFEVEMVASNPGTWLMHCHVDTHVHAGMETLFTVFSR 1068
Qy 343 GALGILKVBGENNHEIYSHKOTDAVYLPBGAPQALDIOEAPRTPAANIOE 393
Db 1069 -----TEHL-----SPLVTYITKETEKP-PRDIEE 1092

RESULT 38
US-10-140-471-234
; Sequence 234, Application US/10140471
; Publication No. US20030138887A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C163
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-471-234

Query Match 4.0%; Score 104; DB 12; Length 1160;
Best Local Similarity 19.7%; Pred. No. 2.5;
Matches 69; Conservative 58; Mismatches 116; Indels 108; Gaps 20;

Qy 102 EYQFWTFGGQVP-----GOMIRVREGDTIEVQFSNHPDSKMPHNVDFHANTG 148
Db 791 EYTDGTF--RIPRPTGPEEHGILGPLKGEVGDILTVFKN--NASRPSVHAH----- 842
Qy 149 PGCGAASFTA-----PGHSTSPFKALQ-----PG-----LYYHCAVAPVGHNIANG 192
Db 843 ---GVLESTTWPLAEPGEVVTYQWNIPEKSGPNDACVSWIYSAVDPI-KDMYSG 898
Qy 193 MYGLI-----LVEKKEGPKVDKEYVW-----QGDFYTKGYGEGQLQPFDMKAIK 240
Db 899 LVGPLAIQCKGILPEHGSRSDMDREFALLFLFDENKSWYLEENAVATGSGQ--DGSINL 956
Qy 241 EDAEVVFNQSVGALTGE-----NALKAKVGETVRLFVNGGPNL---TSSPHVIGELFD 292
Db 957 QDEFLEEN-KHAIINGKLYANRLGLTMYQSERVAMWMLAMGQDVDLHTIHFHASEFLY- 1014
Qy 293 KVHFEKGGENNIOTTLIPAGGAAITEFKVDPGDYVL-----VDH-----AIFRAFNK 342
Db 1015 -----RNGENYRADVDLFPGTFEVEMVASNPGTWLMHCHVDTHVHAGMETLFTVFSR 1068
Qy 343 GALGILKVBGENNHEIYSHKOTDAVYLPBGAPQALDIOEAPRTPAANIOE 393
Db 1069 -----TEHL-----SPLVTYITKETEKP-PRDIEE 1092
```

```
RESULT 39
US-10-140-807-234
; Sequence 234, Application US/10140807
; Publication No. US20030134354A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C174
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-807-234

Query Match 4.0%; Score 104; DB 12; Length 1160;
Best Local Similarity 19.7%; Pred. No. 2.5;
Matches 69; Conservative 58; Mismatches 116; Indels 108; Gaps 20;

Qy 102 EYQFWTFGGQVP-----GOMIRVREGDTIEVQFSNHPDSKMPHNVDFHANTG 148
Db 791 EYTDGTF--RIPRPTGPEEHGILGPLKGEVGDILTVFKN--NASRPSVHAH----- 842
Qy 149 PGCGAASFTA-----PGHSTSPFKALQ-----PG-----LYYHCAVAPVGHNIANG 192
Db 843 ---GVLESTTWPLAEPGEVVTYQWNIPEKSGPNDACVSWIYSAVDPI-KDMYSG 898
Qy 193 MYGLI-----LVEKKEGPKVDKEYVW-----QGDFYTKGYGEGQLQPFDMKAIK 240
Db 899 LVGPLAIQCKGILPEHGSRSDMDREFALLFLFDENKSWYLEENAVATGSGQ--DGSINL 956
Qy 241 EDAEVVFNQSVGALTGE-----NALKAKVGETVRLFVNGGPNL---TSSPHVIGELFD 292
Db 957 QDEFLEEN-KHAIINGKLYANRLGLTMYQSERVAMWMLAMGQDVDLHTIHFHASEFLY- 1014
Qy 293 KVHFEKGGENNIOTTLIPAGGAAITEFKVDPGDYVL-----VDH-----AIFRAFNK 342
Db 1015 -----RNGENYRADVDLFPGTFEVEMVASNPGTWLMHCHVDTHVHAGMETLFTVFSR 1068
Qy 343 GALGILKVBGENNHEIYSHKOTDAVYLPBGAPQALDIOEAPRTPAANIOE 393
Db 1069 -----TEHL-----SPLVTYITKETEKP-PRDIEE 1092

RESULT 40
US-10-140-922-234
; Sequence 234, Application US/10140922
; Publication No. US20030138889A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
```

APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C179
CURRENT APPLICATION NUMBER: US/10/140,922
CURRENT FILING DATE: 2002-05-07
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 234
LENGTH: 1160
TYPE: PRT
ORGANISM: Homo Sapien
US-10-140-922-234

Query*Match 4.0%; Score 104; DB 12; Length 1160;
Best Local Similarity 19.7%; Pred. No. 2.5; Mismatches 116; Indels 108; Gaps 20;
Matches 69; Conservative 58; Mismatches 116; Indels 108; Gaps 20;

QY 102 EYQWTFGGQVP-----GGMIRVREGDTIEVOFSNHPDSKMPHNVDPHAATG 148
DB 791 EYDGTFF--RIPRPTGPEEHLGLPPLKGEVGDITVVFKN--NSRPYSVHAH---- 842
QY 149 PGGAASFTA-----PGHSTFSFKALQ---PG-----LYYHCAVAPVGMHANG 192
DB 843 ---GVLESTVWPLAEPGEVVTYQWNIPEPGRSDSACVSMIYSAVDPI-KDMYSG 898
QY 193 MYGLI-----LYEPKGLPKVDKEYVM-----QGFYTKGKYGEGCLQPPDMKXIR 240
DB 899 LVGPPLAICQKGLPEPHGRSDMDREFALLFLIPDENKSWYLEBNVATHSQ--DPSINL 956
QY 241 EDAYVVFNGSVGALGE-----NALKAKVGETVRLPVNGGPNL---TSSFVIGEIPD 292
DB 957 QDETFLPSN-KMHALINKLYANLRGLMYGGERVAMTLMAGQVDLHTIHFAESFLY- 1014
QY 293 KVAFEGGKGENHNIQTLLIPAGGAATFEKVDVPGDYVL---VDH-----AIPRAFNR 342
DB 1015 -----RNGENYRADVDLFPGTFEVEMVASNPGTWMHCHVDHVGAMETLFTVFSR 1068
QY 343 GALGILKVEGGENHEIYSHKOTDAVYLPBGAPOAIDTOEAPKTPAPANLOE 393
DB 1069 -----TEHL-----SPLTVITKETEKP-PRDIEE 1092

Search completed: August 27, 2003, 18:34:16
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:28:17 ; Search time 29 seconds

(without alignments)
732.416 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626
Sequence: 1 MSKPLIKTTLICALSALML.....NRGGLSADDAVAKKTKPEN 502.

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Issued_Patents_AA.*
- 2: /cgn2_6/prodata/1/1aa/5A_COMB.pep.*
- 3: /cgn2_6/prodata/1/1aa/5B_COMB.pep.*
- 4: /cgn2_6/prodata/1/1aa/5A_COMB.pep.*
- 5: /cgn2_6/prodata/1/1aa/5B_COMB.pep.*
- 6: /cgn2_6/prodata/1/1aa/backfill.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163.5	6.2	645	US-09-252-991A-25253	Sequence 25253, A
2	163	6.2	442	US-09-296-284-26	Sequence 26, App1
3	163	6.2	478	US-09-296-284-5	Sequence 5, App1
4	153	5.8	751	US-09-252-991A-25481	Sequence 25481, A
5	150	5.7	444	US-09-252-991A-17336	Sequence 17336, A
6	150	5.7	712	US-09-252-991A-21481	Sequence 21481, A
7	147.5	5.6	658	US-09-252-991A-24910	Sequence 24910, A
8	146	5.6	468	US-07-985-458-4	Sequence 4, App1
9	146	5.6	593	US-09-252-991A-30673	Sequence 30673, A
10	145	5.5	559	US-08-030-096-6	Sequence 6, App1
11	141	5.4	529	US-08-172-311B-14	Sequence 14, App1
12	141	5.4	529	US-08-706-037-25	Sequence 25, App1
13	141	5.4	529	US-09-005-397-25	Sequence 25, App1
14	141	5.4	529	US-09-032-315-5	Sequence 5, App1
15	141	5.4	529	US-08-993-318A-5	Sequence 5, App1
16	141	5.4	529	US-09-399-886-5	Sequence 5, App1
17	141	5.4	529	US-09-396-260-5	Sequence 5, App1
18	141	5.4	529	US-09-576-281-5	Sequence 5, App1
19	136.5	5.2	326	US-09-615-192A-291	Sequence 291, App
20	136	5.2	320	US-09-252-991A-23061	Sequence 23061, A
21	135	5.1	575	US-09-032-315-8	Sequence 8, App1
22	135	5.1	575	US-08-993-318A-8	Sequence 8, App1
23	135	5.1	575	US-09-399-886-8	Sequence 8, App1
24	135	5.1	575	US-09-396-260-8	Sequence 8, App1
25	135	5.1	575	US-09-576-281-8	Sequence 8, App1
26	133.5	5.1	576	US-08-172-311B-2	Sequence 2, App1
27	132	5.0	671	US-09-328-352-6003	Sequence 6003, Ap

28	131.5	5.0	635	US-09-328-352-8029	Sequence 8029, Ap
29	130	5.0	283	US-08-746-111-5	Sequence 5, App1
30	129	4.9	540	US-08-687-580B-7	Sequence 7, App1
31	125.5	4.8	310	US-09-252-991A-23088	Sequence 23088, A
32	124	4.7	572	US-09-032-315-7	Sequence 7, App1
33	124	4.7	572	US-08-993-318A-7	Sequence 7, App1
34	124	4.7	572	US-09-399-886-7	Sequence 7, App1
35	124	4.7	572	US-09-396-260-7	Sequence 7, App1
36	124	4.7	572	US-09-576-281-7	Sequence 7, App1
37	121.5	4.6	539	US-09-032-315-1	Sequence 1, App1
38	121.5	4.6	539	US-08-993-318A-1	Sequence 1, App1
39	121.5	4.6	539	US-08-689-421-27	Sequence 27, App1
40	121.5	4.6	539	US-09-399-886-1	Sequence 1, App1
41	121.5	4.6	539	US-09-396-260-1	Sequence 1, App1
42	121.5	4.6	539	US-09-389-528-27	Sequence 27, App1
43	121.5	4.6	539	US-09-181-827A-27	Sequence 27, App1
44	121.5	4.6	539	US-09-576-281-1	Sequence 1, App1
45	120.5	4.6	2343	US-09-324-867-2	Sequence 2, App1

ALIGNMENTS

RESULT 1
US-09-252-991A-25253
Sequence 25253, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ IDS: 33142
SEQ ID NO 25253
LENGTH: 645
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25253

Query Match	Score	Length
Best Local Similarity	21.8%, Pred. No. 2.1e-06;	645
Matches	93; Conservative 52; Mismatches 141; Gaps 21;	
QY	19 MMSGCSNQADKA-AOPKSSSTVDAAAKTANADNAASQHGELPVI-----DAIVTHAPEV 72	
DB	41 VLAAGSTPFAKFAAOGGAT-----RGAPVLKGTDFDVIASPV- 81	
QY	73 PPVVDHDAKAVVVKMETVEKVMRLADGVEYQFWTFGGVPCQMIRVREGDTIEVQSNH 132	
DB	82 -----NFGKRGVAT-----TINGSLPAPTLRMWRGDFVITVTN- 116	
QY	133 PDSKMPHNDFFAALGP---GGGAASFT--APGHSTSPFALQGLVYHCAVAPVM 187	
DB	117 -KLEKRTSITHMGIILPYOMDGPVGISFNGIAPGETFYRFVQSGSYWH---SHSGF 172	
QY	186 HIANGVGLIPEKGLPKV---DKEYVYM-----OGDFYT----- 221	
DB	173 QELTGMGALITIDP-AGATITADRDHVLPSFDMDDEDPMRFYTKLSQSDYVNVQPTV 231	
QY	222 -----KGKYGEOGLQPFDMKAIREDAAVEVNGSVGALTGENALKA 263	
DB	232 PFFRDASREGTAALIDKKKMNEMKMNPTDLADISATLITVLA--NGVTPA--GWTALF 288	
QY	264 KXGEVRLPVGNGGNLTSFHVIGEIPKXVPEGGKGNHNIQTTLIAGAGAITERFV 323	
DB	289 RGERVRLMVGAGNTFYDVRIPLGLKLTVAHVG-----V 324	

Qy	334	DVPGGVVLVDNHI	IFRAFNKG	-ALGILKVEEENHEIYSHKQTAVVLP	ECAPAOIPDQEA	382
Db	325	DV--SPVTVDE--FR	-FGPGETVDVIVQPRDDAVTI	FAQAMDITGV----	ARATLAVREG	375
Qy	383	PKTPAPA	369			
Db	376	LOAPVPA	382			

RESULT 2
US-09-296-284-26
; Sequence 26, Application US/09296284A
Date: 09/20/2010

: TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
 : TITLE OF INVENTION: and Methods of Use Thereof
 : FILE REFERENCE: 1533.0870000
 : CURRENT APPLICATION NUMBER: US/09/296,284A
 : CURRENT FILING DATE: 1999-04-22

```

; SEQ ID NO 26
;
; LENGTH: 442
;
; TYPE: PRF
;
; ORGANISM: Gluconobacter suboxydans
;
US-09-296-284-26

```

Query Match	6.2%	Score 163;	DB 3;	Length 442;
Best Local Similarity	37.3%;	Pred	No. 1.2e-06;	
Matches 50;	Conservative 17;	Mismatches 57;	Indels 10;	Gaps 6

369 LPEGAQAIDTQEAPTPAPANLQEQIKAGKATYDSCAACQPDGKVPNAPPLNSD 428
 270 VPEGKTLG---QDDGKATALLGAGKGADAGAEVYLHNCAICMNDGCGVNRNFPPLGAP 326

Db

```
429 YLNMDHARA-ASIVANGLSGKITVNGNQYESI-MPALA--ISDQOIANVIYTLNLSFGNK 4884  
429 :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
327 VVINDNATSMANIVTFG--GLIPPTTAPASAVAMPGRDHLSDDQILDVDVNFMKRSWGNO 3844
```

```

OY      485 G-GQLSADDVAKAK 497
        |||||:|:
Db      385 APGTLASDIRKLR 398

```

RESULT 3
US-09-296-284-5
; Sequence 5, Application US/09296284A

; APPLICANT: Choi, Eun-Sung
 ; APPLICANT: Rhee, Sang-Ki
 ; APPLICANT: Lee, Eun-Hae
 ; TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes
 ; TITLE OF INVENTION: and Methods of Use Thereof

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 478
; TYPE: PR1
; ORGANISM: Gluconobacter suboxydans
; US-09-236-284.5

```

Query Match 6.2%; Score 163; DB 3; Length 478;
 Best Local Similarity 37.3%; Pred. No. 1.4e-06;
 Matches 50; Conservative 17; Mismatches 57; Indels 10; Gaps 6

Db 306 VPEGKULG---QDDGKATALLAEGGKGAGAEVILHNCALCHMNDGTVGVMRMEFPLLAGNP 362

Qy 429 YLNADHARA-ASTIVANGLSGKITTNGQYESV-APALA--LSDOIINVITYLTNSGCK 484

Db 363 VATTDAISMALVTFG--GIUPTNTAPSAVAMFGFDHLSDDQIADVVMFKMSGNQ 420

```
QY      485 G-GQLSADDDVAKAK 497
        | | | | : | :
Db      421 APGTLASDIRKLR 434
```

RESULT 4
US-09-252-991A-25481

: APPLICANT: Marc J. Rubenfield et al.
 : TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
 : TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: 107196.136
 : CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 25481

```

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:

```

```

: LOCATION: (339)
: OTHER INFORMATION: Identity of amino acid at the above locations are unknown
US-09-252-991A-25481

```

Query Match 5.8%; Score 153; DB 4; Length 751;
 Best Local Similarity 35.5%; Pred. No. 2.5e-05;
 Matches 38; Conservative 15; Mismatches 48; Indels 6; Gaps 4

QY 396 KAGATATYDNCACHQPDGCGVPNAFPPLA-NSDYLINADHARAASIVANGLSGKITVNGN 454
| | | | | : | | | | | :
Db 626 KTGAALYYDNCAGCHRTDQKGYARVEPALAGNPVVTGSDPTSLVHIVLKG--GTLPLATHQ 6833

```

QY      455 QYES--VMPALA--LSDQIANVITYTLNSFGNKGGLSADDVAKAKK 4988
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      684 APSSFTMPFGMRMNDQEIADVNFIRTSWGNQAPSVSVDVRRRLRK 7300

```

RESULT 5
US-09-252-991A-17336
; Sequence 17336, Application US/09252991A

APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,180
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 17336

ORGANISM: *Pseudomonas aeruginosa*
US-09-252-991A-17336

SOFTWARE: ASCII Form
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/985,458
FILING DATE: 19921203
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/658,221
FILING DATE: 20-FEB-1991
APPLICATION NUMBER: 73440/1990
FILING DATE: 26-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Goodman, Herbert
REGISTRATION NUMBER: 17081
REFERENCE/DOCKET NUMBER: 910134/HG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)972-1400
TELEFAX: (212)370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: MATURE PEPTIDE
LOCATION: 24 to 468
IDENTIFICATION METHOD: similarity of other signal
IDENTIFICATION METHOD: sequences
ORIGINAL SOURCE:
ORGANISM: Acetobacter alioacetigenes
STRAIN: MH-24
PUBLICATION INFORMATION:
AUTHORS: Tamaki, Toshimi;
AUTHORS: Fukaya, Masahiro;
AUTHORS: Takemura, Hiroshi;
AUTHORS: Takemura, Kenji;
AUTHORS: Okumura, Hajime;
AUTHORS: Kawamura, Yoshiya;
AUTHORS: Nishiyama, Makoto;
AUTHORS: Horinouchi, Sueharu and
AUTHORS: Bepko, Teruhiko
TITLE: Cloning and sequencing of the Gene Cluster
TITLE: Encoding Two Subunits of Membrane-Bound
TITLE: Alcohol Dehydrogenase from Acetobacter
TITLE: polyoxenes
JOURNAL: Biochimica et Biophysica Acta.
VOLUME: 1088
PAGES: 292-300
DATE: 1991
US-07-985-458-4

Query Match 5.6%; Score 146; DB 1; Length 468;
Best Local Similarity 25.2%; Pred. No. 5.2e-05;
Matches 84; Conservative 50; Mismatches 127; Indels 72; Gaps 19;

184 PVGHNINAGVGLL-VERKELRKYDKEYVVGSGDYTKGKGEQGL- -QPFQV-EK 237
149 PLSNRWPLGIMRMKSPSPKDFTRPAGTDEIARGLVLTGP-CHCACHTPKGFAMQEK 207
238 AIRDEAEVYVNGSVGALTEGNALAKAVGETVR- -LFVNGGPNLTSSFHVIGEIFDK 293
208 AL- -DA- -GGPDLGSGAPRIDNVVAPSLRNDPVVGLRWSEDDIYTFILSGRIDHS 260
294 VHEEGGGENHNIOTTLIPAGGAAITFPKYDVEDYLV- -HAIFPAFKGALGILKVE 351
261 AVF-GGWD- - - - -VWAMSTQYFTDDLAIAK-YLKSILRPVPSQ 299
352 GEENHEIYSHKQTDVAVLPFGARQALDQEARPKTPRANLOEQKAKATYDSNCAACHQ 411
300 GN- - - - -YTDPSFANPLASG- - - - -NTASV- - - - -GADTYVKECALCH 335
412 PDGKGVNAPPLANSYLNADHARA-ASIVANGLSKITVNGQYESV-MPAI- -ALSD 467

336 NDGGVARNMPPLAGPVVVTENPTSLVNVIAHG- -GVLPSSWABSAVAMPQSKSLA 393
468 QQIANVITYTLNPFNGK- -GQLSADDAVAKAKT 499
394 QQIADVNFRTISWGNKAPGTVAADVTGLRDT 426

RESULT 9
US-09-252-991A-30673
Sequence 30673; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubinfeld et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30673
LENGTH: 593
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30673

Query Match 5.6%; Score 146; DB 4; Length 593;
Best Local Similarity 22.1%; Pred. No. 7.7e-05;
Matches 116; Conservative 61; Mismatches 219; Indels 128; Gaps 25;

68 HAEVPPVYDRDRAKVVVYKMETVEKMRADGVEYQFWTFGGQV- -QMRVREGDT 124
92 HPRAEPPAEVRRPVLV- - - - -LRAGLRORRRRAGGKSPAPADQORAGADR 143
125 IEVQFSNHPDSKMPHNVDFH- - - - -AATGPG-GGAASFPAPGHTSFSPKALOP- - - - - 173
144 GQPRSHLPLHRLRALLQNHPRARTRRPGGGLRKRKTLSC- -LALALAAISPFLASAA 201
174 - - - - -GLVYH- - - - -CAVAPVGNHANGM- - - - -YGLILVEPK- - - - -GDP 206
202 GADALVKRGEYLARADACHTAGAPFAGGLIQSPFTGYGTNTTPKHEGIGAY 261
207 KYDKETVYVNGDYTYTKGYEQQ- - - - -LOPFMEKAI-REDAEYV- - - - - 247
262 SADEFPAL- - - - -TEGKRKDGAVLYVPAMPYTSYHLIEREDADAIYAVLMAQEPILRAP 316
248 - - - - -FNGSVGALTEGNALAKAVGETVR-LFVNGGPNLTSSFHVIGEIFDKVHREGK 300
317 QTSLSFPPVNRNG-LAGMWL- - - - -YKSVLRLOPBERGSRAMKRGQYWEVL- - - - -GHC 366
301 GENHNIOTTLIPAGGAITEFKVD- - - - -VPGD- - - - -YVLVDHAIIFRAFNG 343
367 GECHTFRN- - - - -LAGLEDDKRLSGSLNGYLARSLRPDLARGTQGLDLSFLHGMS 422
344 ALGIL- - - - -KVEGENHEIYSHKQTDV- - - - -LPFGARQALDQEARPKTPRANLOEQIK 396
423 AQSMEFENFPPVNHSTQHLDDSLAAMATYLLGDPQPPAKVQALFE- - - - -AQLNDSG 478
397 AGKATVDSNCAACHQPDGKGVN- - - - -APPLANSDYLNADHARAASIVANGLSKITVNGNO 455
479 RGRQOVLNACAGHGSGGKPRHIAVAMNGTTLRLQDRPNLIRVEDGIVBO-QPTGFE 537
456 YESVMPAIA- - - - -LSDOQIANVITYTLNPFNGKSGQLSADDAVAK 497
538 RMQPMGFGAKLDDDEGLTDLNLYLRTQWGLRGFDLGRQVADLK 581

RESULT 10
US-08-030-096-6

; Sequence 6, Application US/08030096
; Patent No. 5426041
; GENERAL INFORMATION:
; APPLICANT: Fabianski, Steven F.
; APPLICANT: Arison, Paul G.
; TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID
; TITLE OF INVENTION: SEED PRODUCTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/030,096
; FILING DATE: 22-MAR-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/556,917
; FILING DATE: 20-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA91/00255
; FILING DATE: 22-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/164/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5390
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 559 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-030-096-6

Query Match 5.5%; Score 145; DB 1; Length 559;
Best Local Similarity 26.1%; Pred. No. 8.7e-05;
Matches 71; Conservative 37; Mismatches 132; Indels 32; Gaps 12;
QY 100 GVEYQFWTFGGQVPGQMIRVBEGDTIEVQFSNHPD-----SKMPHNVDFHAATGPGGG 152
DB 43 GVPQVILINGQFPGRNINSTNNNNIIVFNNLDEPFLITNNGIGHKNCMODGTPTM 102
QY 153 ABASTAPGHSTFSFKAL-QPGLYYHCAVAFVGMHIANGMGLILVEPKGLP---K 207
DB 103 CP---IMPETNYTHYHPOPKDQIGSYFY---PTTGMHRAAGYGGGRVNSRLIIPVYAD 156
QY 208 VDEKEYVMGDEYTTKKGVEGGLQPF-DHEKAI-REDAEYVVPNGSVGLTGENA--LKA 263
DB 157 PDDDTVLIGDMYTK---SHTQIKKPLDGRRTIGRPDG---IYINGSGKDGSDALEFTL 211
QY 264 KVGETVRLFEVNGGPNLTSSFVHIGEIFDKVHEGEGKGENHNIOITLLPAG---GAITE 320
DB 212 KKGKTYRVAICVGVTSINFRIOHNMKMLVEMESHVONDIYSDLVAVGCGFGITVTA 271
QY 321 FKVDVPGDVLVDHAIERAFNKGALGILKEVG 352
DB 272 NO--EPKDYVMVASSRFLKTVITTTGTLRYEG 301

RESULT 11
US-08-172-331B-14
; Sequence 14, Application US/08172331B

; Patent No. 5480801
; GENERAL INFORMATION:
; APPLICANT: Wahlethner, Jill A.
; APPLICANT: Christensen, Bjoern E.
; APPLICANT: Schneider, Palis
; TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5480801 of No. 5480801 of No. 5480801 of America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/172,331B
; FILING DATE: 22-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/122,230
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/122,827
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/162,827
; FILING DATE: 03-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4052.020-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-172-331B-14

Query Match 5.4%; Score 141; DB 1; Length 529;
Best Local Similarity 20.4%; Pred. No. 0.00019;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;
QY 59 LPVIDAIVTHADVEPPVDRDHPAKVVK--METVEKVRNLADGVYQFWTFGGQVPGQM 116
DB 8 LPLLAVST-----PAPAFAVNYKFDIKNVAVADGQRSIVSVNGLVPCTL 54
QY 117 IVRREGDTIEVQFSNH---PDSKMPHNVDFH---AATPGGGAESTFAP-----GHS 164
DB 55 ITANKGDTLRINVTNQLDPSNRKATTTIHGQLFOATTADEDPAPFVTCPLAQNLSYTY 114
QY 165 TFSFALPGLLVYHCAVAPVGMHIANGMVGL-LILVEP---KEGLPKVDEKEYVMGDP 219
DB 115 EIRPLAG-QGTIMWYH---AHLASQVVDGLRGVLVIDPRDPKRSYDVDDASTVNLLEW 170
QY 220 Y-----TKGKYGEGQ---IQPFMEKAIREDAYVVPNGSVGLTGENALKARVGETV 269
DB 171 YHTPAVLEKQMFSTNNNTALLSPVPDGLINSGRYV-----GDAVPRSVINVKRGKRY 225
QY 270 RLFEVNGGPNLTSSFVHIGEIFDKVHEGEGKGENHNIOITLLPAG---AATIEFKVD 324
DB 226 RLRVINASAGISFTSY-----EGHSL--TVIEADGILHQLPLAVDSFOLY 268
QY 325 VPGDVLVDHAIERAFNKGALGILKEVEGENHEIYSHKOTDAVVPYEPGAQAIDTQE--- 381

Db 269 AGORYSVIYEANOTANWIRAPMTVAGAGTNANLDPFTNVFAVLHGYGADNAPTTEQGS 328
QY 382 -----AKTPAPANLQOIKAGKATYDS-----NCAACHOPQKGV 417
Db 329 AIGTALVEENLHALINPGAPGSAADVSLNLAIGRSTVDGILRFTFNNIKYEAP----- 383
QY 418 PNAFPPPLANSDYLNADHARAASIVANGLSGKITVNGQYESVMPALALSDQOIANVITYT 477
Db 384 --SLPTL-----LKLANNASNDADFTNEHTIVLPHNKVIELNITGADHP 428
QY 478 LNSFGKGGQLSADDAVAKAKTKPN 502
Db 429 IHLHG-----VFIDVKSLSGTPN 447

RESULT 12

US-08-706-037-25
Sequence 25, Application US/08706037
Patent No. 5770419
GENERAL INFORMATION:
APPLICANT: Xu, Feng
APPLICANT: Berka, Randy M.
APPLICANT: Wahlechner, Jill A.
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 57704199 No. 5770419disk of No. 5770419th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706.037
FILING DATE: 30-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,800
FILING DATE: 1-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4526.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-706-037-25

Query Match 5.4%; Score 141; DB 1; Length 529;
Best Local Similarity 20.4%; Pred. No. 0.00019;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;
QY 59 LPVIDAIVTAPEVPPVDRDHPAKVVK--METVEKVMRLADGVEYQFTFGQVPGQM 116
Db 8 LPLAAVST-----PAPAVRNVKFKFDIKVNVAPDGFQSRISVNGLVPGTL 54
QY 117 IRRREGTIVQPSNH---PDSKMPHNVDH---AATGGCGGAESAFTAP-----GHTS 164
Db 55 ITRAKGDTLRINVTNQLDTPSRMRATTIHHGLFQATTADEDDGPAFVTCQPIAONISITY 114

QY 165 TFSFKALQGLVYVHCAPVPGNHIANMGY-LILVEP---KEGLPKYDKEYVYVQSGF 219
Db 115 EIPLRG-QGTMMVYH---AHLASQYVDGLRGLPLVYDPPNDPKSRVYDADASTVWVLEDM 170
QY 220 Y-----TKGKGEQG---LOPFMEKAIREDAEYVNVNGSVGALTGENALKAKGVEY 269
Db 171 YHTPAFVLEKQMTSTNTTALLSPVDSGLINGKGRVY-----GGPVPVRSVINVRKRGKY 225
QY 270 RLPVNGGPNLNTSPFVIGEIPDKVHFBGKGGENHNIQTLLPAGS-----AATFEKVD 324
Db 226 RLKVTINASAIGSTFETI-----EGHSL--TVLEADSLIQLPLAVDSFOIY 268
QY 325 VPQDYVLVDHAIFFAPFNKALGILKYGEENHEIYSHKOTDAVYLEPAGQALDQGE--- 381
Db 269 AGORYSVIYEANOTANWIRAPMTVAGAGTNANLDPFTNVFAVLHGYGADNAPTTEQGS 328
QY 382 -----AKTPAPANLQOIKAGKATYDS-----NCAACHOPQKGV 417
Db 329 AIGTALVEENLHALINPGAPGSAADVSLNLAIGRSTVDGILRFTFNNIKYEAP----- 383
QY 418 PNAFPPPLANSDYLNADHARAASIVANGLSGKITVNGQYESVMPALALSDQOIANVITYT 477
Db 384 --SLPTL-----LKLANNASNDADFTNEHTIVLPHNKVIELNITGADHP 428
QY 478 LNSFGKGGQLSADDAVAKAKTKPN 502
Db 429 IHLHG-----VFIDVKSLSGTPN 447

RESULT 13

US-09-005-397-25
Sequence 25, Application US/09005397
Patent No. 5972670
GENERAL INFORMATION:
APPLICANT: Xu, Feng
APPLICANT: Berka, Randy M.
APPLICANT: Wahlechner, Jill A.
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 59726700 No. 5972670disk of No. 5972670th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,397
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/706,037
FILING DATE: 30-AUG-1996
APPLICATION NUMBER: US 60/002,800
FILING DATE: 1-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4526.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-005-397-25

Query Match
Best Local Similarity 20.4%; Score 141; DB 2; Length 529;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

59 LPVIDAIVTHAPEVPPVDRDHPAKVVK--METVEKVMRLADGVGYQFTFGGVPQGM 116
8 LPLLAAVST-----PAPAANRYKFKDKNVAVAPDGFORSIVSVGLVPGTL 54
117 IAVREDDTIEVQFSNH---PDSKMPHNVDPH---AATGGGAASFTAP-----GHTS 164
55 ITANKGDTLRINTVNTQLTDPMSMRATTIHHGLFOATTDEDPARVTOCPFIQNLSTYT 114
165 TFSFKALQGLVYVHCAPVGMHIANGVG-LIIVP-----KEGLPKVDKEYVYVNGQDF 219
115 EIPLRG-QGTVMVYH---AHLASQYVDGLRGLVIVDPNDPKSRVYDVDDASTVVMLEDM 170
220 Y-----TKGKYGEG---LQPFDMEKAIREDAEVYVNGSVGALTGENALAKVGETV 269
171 YHTPAVLEKQMFSTNTNLTLLSPVPSGLINGKGRVY-----GGPAVPSVIVNKKGRY 225
270 RLFPVNGGPNLTSSFVIGEIFDKVHEGSGKENHNIQTLLPAG-----AAITEPKVD 324
226 RLKRVINASAGSFTFSI-----EGHSL--TVIADGILHQLPLAVDSFOIY 268
325 VEGDYLVDHAI FRAFNKGLGILKVEGEEHNIYSHKOTDAVYLPBGAPOALDIOE--- 381
269 AGORYSVIVEANQTAANYIRAPMTVAGACTNANLDPFTVFAVLHYEGAPNAEPTTEQGS 328
382 -----APKTPAPANLQEQIKAGATYDS-----NCAACHQPDGKV 417
329 AIGTALVEENLHALINPGAPGSAPADVSLNLAIGSTVDGILRFTFNNIKYEAP----- 383
418 PNAFPPLANSDYLNADHARAASIVANGSGKITVNGNOYESVPAIALSDQOIANVITYT 477
384 --SLPFTL-----LKLANNASNDADFTPNEHTIVLPHNKKVIELNITGGADHP 428
478 LNSFGKGGQLSADDAVAKAKTKPEN 502
429 IHLHGH-----VFDIVKSLGGTPN 447

RESULT 14
Sequence 5, Application US/09032315

Patent No. 5985818
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 59858180 No. 5985818disk of No. 5985818th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,315
FILING DATE: 27-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5200.200-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-032-315-5

Query Match
Best Local Similarity 20.4%; Score 141; DB 2; Length 529;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

59 LPVIDAIVTHAPEVPPVDRDHPAKVVK--METVEKVMRLADGVGYQFTFGGVPQGM 116
8 LPLLAAVST-----PAPAANRYKFKDKNVAVAPDGFORSIVSVGLVPGTL 54
117 IAVREDDTIEVQFSNH---PDSKMPHNVDPH---AATGGGAASFTAP-----GHTS 164
55 ITANKGDTLRINTVNTQLTDPMSMRATTIHHGLFOATTDEDPARVTOCPFIQNLSTYT 114
165 TFSFKALQGLVYVHCAPVGMHIANGVG-LIIVP-----KEGLPKVDKEYVYVNGQDF 219
115 EIPLRG-QGTVMVYH---AHLASQYVDGLRGLVIVDPNDPKSRVYDVDDASTVVMLEDM 170
220 Y-----TKGKYGEG---LQPFDMEKAIREDAEVYVNGSVGALTGENALAKVGETV 269
171 YHTPAVLEKQMFSTNTNLTLLSPVPSGLINGKGRVY-----GGPAVPSVIVNKKGRY 225
270 RLFPVNGGPNLTSSFVIGEIFDKVHEGSGKENHNIQTLLPAG-----AAITEPKVD 324
226 RLKRVINASAGSFTFSI-----EGHSL--TVIADGILHQLPLAVDSFOIY 268
325 VEGDYLVDHAI FRAFNKGLGILKVEGEEHNIYSHKOTDAVYLPBGAPOALDIOE--- 381
269 AGORYSVIVEANQTAANYIRAPMTVAGACTNANLDPFTVFAVLHYEGAPNAEPTTEQGS 328
382 -----APKTPAPANLQEQIKAGATYDS-----NCAACHQPDGKV 417
329 AIGTALVEENLHALINPGAPGSAPADVSLNLAIGSTVDGILRFTFNNIKYEAP----- 383
418 PNAFPPLANSDYLNADHARAASIVANGSGKITVNGNOYESVPAIALSDQOIANVITYT 477
384 --SLPFTL-----LKLANNASNDADFTPNEHTIVLPHNKKVIELNITGGADHP 428
478 LNSFGKGGQLSADDAVAKAKTKPEN 502
429 IHLHGH-----VFDIVKSLGGTPN 447

RESULT 15
US-08-993-318A-5
Sequence 5, Application US/08993318A

Patent No. 5998353
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Palle
APPLICANT: Raemussen, Grethe
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 59983530 No. 5998353disk of No. 5998353th America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/993,318A
 FILING DATE: December 18, 1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Gregg, Valeta A.
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 5032.200-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 529 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-993-318A-5

Query Match 5.4%; Score 141; DB 2; Length 529;
 Best Local Similarity 20.4%; Pred. No. 0.00019;
 Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPVPPVVDHDPKAVYK--MEIVEKWLADGVEYQFTFGQVPGOM 116
 DB 8 LPLLAAVST-----PAPAFAVNVKFDIKVNVAPDPGFSRISVNGVLPGTL 54
 QY 117 IRVREGDTIEVOFSNH---PDSKMPHNVDFH---AATGGGAASFTAP-----GHTS 164
 DB 55 ITANKGDTLRINVTNQLTDPMSKRATTIMHGLFQATTADEDEGPAFTVOCPIAONLSYTY 114
 QY 165 TFSFKALQPLGLVYHCAVAPVGMHIANMGY-LILVEP---KEGLPKYDKXYVNGQGF 219
 DB 115 EIPLRG-QGTGMWYH---AHLASQYVDGLRGPLVIYDPNDPKSRVYVDVASTVWLEDM 170
 QY 220 Y-----TKGKGEQG---LQPFDMKAIREDAEYVFNQSGVALTGENALAKVGETV 269
 DB 171 YHTPAPVLEKQMFSTNTTALLSPVDSGLNGKGRVY-----GGPAVPRSVINVKGRKY 225
 QY 270 RLFVNGGPNLTSSFFHVGIEIFDKVFEKGGENHNIQTTLLIPAG-----AATPEKVD 324
 DB 226 RLKVINASAGSFTFSI-----EGHSL--TVIEADGILHQPPLAVDSFQIY 268
 QY 325 VPQGDYLVDAIFRAFKGALGILKVEGGENHEIYSHKQTDVAVYLPQGAPOAIDTQE--- 381
 DB 269 AGQRYSVIVEANQTAANYWIRAPMTVAGAGTNANLDPNTVFAVLHYEGAPNAEPTTEQGS 328
 QY 382 -----AKTPAPANLQEOIKAGKATYDS-----NCAACHQPDGKV 417
 DB 329 AIGTALVEENLHALINPGAGSAPADVSINLAIGSTVDGILRTFNNIKYAP----- 383
 QY 418 PNAFPPLANSDYLNADHARAASIYANGLSGITVNGQYESVMPAIALSDQOIANVITYT 477
 DB 384 --SIPTL-----LKIILANNASNDADTPMEHTITVLPHNKVIELNITGGADHP 428
 QY 478 LNSFGNKGQLSADDAVAKAKTKPN 502
 DB 429 IHLHG-----VPDIVKSLGCTGN 447

RESULT 16
 US-09-399-886-5
 Sequence 5, Application US/09399886
 Patent No. 6140092
 GENERAL INFORMATION:
 APPLICANT: Pedersen, Anders
 APPLICANT: Svendsen, Allan
 APPLICANT: Schneider, Paule

APPLICANT: Raasmussen, Grethe
 APPLICANT: Cherry, Joel
 TITLE OF INVENTION: LACCASE MUTANTS
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSER: No. 61400920 No. 6140092disk of No. 6140092ch America
 STREET: 405 Lexington Avenue
 CITY: New York
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/399,886
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/993,318
 FILING DATE: December 18, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Gregg, Valeta A.
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 5032.200-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 529 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-399-886-5

Query Match 5.4%; Score 141; DB 3; Length 529;
 Best Local Similarity 20.4%; Pred. No. 0.00019;
 Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPVPPVVDHDPKAVYK--MEIVEKWLADGVEYQFTFGQVPGOM 116
 DB 8 LPLLAAVST-----PAPAFAVNVKFDIKVNVAPDPGFSRISVNGVLPGTL 54
 QY 117 IRVREGDTIEVOFSNH---PDSKMPHNVDFH---AATGGGAASFTAP-----GHTS 164
 DB 55 ITANKGDTLRINVTNQLTDPMSKRATTIMHGLFQATTADEDEGPAFTVOCPIAONLSYTY 114
 QY 165 TFSFKALQPLGLVYHCAVAPVGMHIANMGY-LILVEP---KEGLPKYDKXYVNGQGF 219
 DB 115 EIPLRG-QGTGMWYH---AHLASQYVDGLRGPLVIYDPNDPKSRVYVDVASTVWLEDM 170
 QY 220 Y-----TKGKGEQG---LQPFDMKAIREDAEYVFNQSGVALTGENALAKVGETV 269
 DB 171 YHTPAPVLEKQMFSTNTTALLSPVDSGLNGKGRVY-----GGPAVPRSVINVKGRKY 225
 QY 270 RLFVNGGPNLTSSFFHVGIEIFDKVFEKGGENHNIQTTLLIPAG-----AATPEKVD 324
 DB 226 RLKVINASAGSFTFSI-----EGHSL--TVIEADGILHQPPLAVDSFQIY 268
 QY 325 VPQGDYLVDAIFRAFKGALGILKVEGGENHEIYSHKQTDVAVYLPQGAPOAIDTQE--- 381
 DB 269 AGQRYSVIVEANQTAANYWIRAPMTVAGAGTNANLDPNTVFAVLHYEGAPNAEPTTEQGS 328
 QY 382 -----AKTPAPANLQEOIKAGKATYDS-----NCAACHQPDGKV 417
 DB 329 AIGTALVEENLHALINPGAGSAPADVSINLAIGSTVDGILRTFNNIKYAP----- 383
 QY 418 PNAFPPLANSDYLNADHARAASIYANGLSGITVNGQYESVMPAIALSDQOIANVITYT 477

Db 384 --SLPTL-----LKLANNASNDADFTPNEHTTVLPHNKVIELNTGGADHP 428

QY 478 LNSFGNGQLSADYAKAKTKPEN 502
: : : : :
Db 429 IHLHG-----VFIVKSLGGTNP 447

RESULT 17
US-09-396-260-5
; Sequence 5, Application US/09396260
; Patent No. 6184015
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Xu, Feng
; TITLE OF INVENTION: LACCASE MUTANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 61840150 No. 6184015disk of No. 6184015th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/396, 260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,315
; FILING DATE: 27-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5200.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; *LENGTH: 529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-396-260-5

Query Match 5.4%; Score 141; DB 3; Length 529;
Best Local Similarity 20.4%; Pred. No. 0.00019;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPVPPVDRDHPAKVVK--METVEKVMRLADGVEYQFTFGGQVPGQM 116
: : : : :
Db 8 LPLLAIVST-----PAPFAVNNYKFDIKGVNVVAPDGFORSIVSVGLVPGTL 54

QY 117 IRVREGDTLEVQPSNR---PDSKMPHNVDPH---AATGPGGGAESFTAP-----GHTS 164
: : : : :
Db 55 IFANKGDTLRINTVNTQUTDPSMRATYIHHGAFQATTADEDPAPFTQCPRIQNSITY 114

QY 165 TFSFKALQGLLYYHCAVAPVGHIANMGK-LILVER---KEGLPKVDKEYVYVMOGDF 219
: : : : :
Db 115 EIPLRG-QGTGMWYH---AHLASQYVDGLRGLPLVIDPNDPHKSRVDDVASTVVMLEW 170

QY 220 Y-----TRKGYGEOQ---LOPFDMEKAIRDEAEYVFNFGVGLTGENALAKVGETV 269
: : : : :
Db 171 YHTRPAPVLEKOMSTNTALLSPVDSGLINGKRYV-----GPAVPRSVIVNKKGRY 225

QY 270 RLTFVNGGPNLTSSFHVIGEIFPKVHEGKGGENHNIQTLLIPAG-----AAITEPKVD 324
: : : : :
Db 8 LPLLAIVST-----PAPFAVNNYKFDIKGVNVVAPDGFORSIVSVGLVPGTL 54

Db 226 RLRFVNASAIIGSTFSI-----EGHSL--TVIEADGILHQPALVDSFOIY 268

QY 325 VPGDYLVDHALFRAPFKALGILKVEGENHEIYSHKOTDAVYLPREGAPOAIDTOE--- 381
: : : : :
Db 269 AGGRSVIVIBANQNTANWYIRAPMTVAGAGTANNDPTVFPVFLHIEGAPNMEPTTEQS 328

QY 382 -----APKTPAPANLQEOIKAGATYDS-----NCAACHQPDGKV 417
: : : : :
Db 329 AIGTALVEENLHALINPGAPGSGAPADVSLNLAIGSTVDGILRFTEFNNIKYEP----- 383

QY 418 PNAFPLPANSIDYINADHAPAAISVANGLSGKITVNGNOYESWPAIALSDQOIANITYT 477
: : : : :
Db 384 --SLPTL-----LKLANNASNDADFTPNEHTTVLPHNKVIELNTGGADHP 428

QY 478 LNSFGNGQLSADYAKAKTKPEN 502
: : : : :
Db 429 IHLHG-----VFIVKSLGGTNP 447

RESULT 18
US-09-576-281-5
; Sequence 5, Application US/09576281
; Patent No. 627611
; GENERAL INFORMATION:
; APPLICANT: Pedersen, Anders
; APPLICANT: Svendsen, Allan
; APPLICANT: Schneider, Palle
; APPLICANT: Rasmussen, Gretie
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: LACCASE MUTANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6276110 No. 627611disk of No. 627611th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/576, 281
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,318
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valera A.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 5032.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; *LENGTH: 529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-576-281-5

Query Match 5.4%; Score 141; DB 3; Length 529;
Best Local Similarity 20.4%; Pred. No. 0.00019;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPVPPVDRDHPAKVVK--METVEKVMRLADGVEYQFTFGGQVPGQM 116
: : : : :
Db 8 LPLLAIVST-----PAPFAVNNYKFDIKGVNVVAPDGFORSIVSVGLVPGTL 54

QY 117 IRVAGDTIEVQFSNH---PDSKMPHNVDFH---AATGPGGAASFTAP-----GHTS 164
 Db 55 ITANKGDTLRINVTNQLDPSMRRAITIIHMGLEQATTADEBDGAFYTOCPDIANLSYTY 114
 QY 165 TFSFKALQPGILYTHCAVAVGMHIANWNG-LIVER---KEGLPKVDKEYVWQGF 219
 Db 115 EIPLRG-QGTGMWH---AHLASQYVGLRGPVLYIPNDPKRKYVDVDASTVWMLDW 170
 QY 220 Y-----TGKVEGQ---LQPDMEKAIREDAEVVFNGSVALLGENAKKAGETV 269
 Db 171 YHPPAPVLEKOMFSTNTTALISVPDGLNGKGRY---GSPAVPRSVINVKRKY 225
 QY 270 RLFGVNGSPULTSFHVIGEIFDKHFEKGGENHNIQTLLIPAGG-----AATFEKVD 324
 Db 226 RLAVINSAISGFTFSI-----EGHSL--IYIADGILHQPLAVDSFOY 268
 QY 325 VPGDYVLVDHAIFFAPFKKALGILKVEGENHEIYSHKQTDVYLLPGAPQALDTOR--- 381
 Db 269 AGQRYSVIIVEMQTPAAVWTRAPMTVAGAGTNNALDPTNVFAVLHVGAPNAEPTTEGGS 328
 QY 382 -----APKTPANLQEQIKAGKATYDS-----NCAACHQPDGKV 417
 Db 329 AIGTALVEENLHALINPGAPGSPADVSLNLAIGRSTVDGILRFTFNNIKYEAP----- 383
 QY 418 PNAFPPLANSDYLNADHARAASIVANGLSGKITVNGQVESVMPAIALSDQIANTVITYT 477
 Db 384 --SLFTL-----LKIANNASNDADPTPEHNTIYVLPNNVIELINTIGADHP 428
 QY 478 LNSFGKGGQLSADDAVAKAKTKEN 502
 Db 429 IHLHG-----VFDIKSLGCTPN 447

RESULT 19
 US-09-615-192A-291
 ; Sequence 291, Application US/09615192A
 ; Patent No. 6410718
 ; GENERAL INFORMATION:
 ; APPLICANT: Blokeberg, Leonard N.
 ; TITLE OF INVENTION: Materials and Methods for the
 ; FILE REFERENCE: 11000.1003C4U
 ; CURRENT APPLICATION NUMBER: US/09/615,192A
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 08/975,316
 ; PRIOR FILING DATE: 1997-11-21
 ; PRIOR APPLICATION NUMBER: US 08/713,000
 ; PRIOR FILING DATE: 1996-09-11
 ; PRIOR APPLICATION NUMBER: US 09/169,789
 ; PRIOR FILING DATE: 1998-10-09
 ; NUMBER OF SEQ ID NOS: 405
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 291
 ; LENGTH: 326
 ; TYPE: PRT
 ; ORGANISM: Eucalyptus grandis
 US-09-615-192A-291

Query Match 5.2%; Score 136.5; DB 4; Length 326;
 Best Local Similarity 20.9%; Pred. No. 0.00022;
 Matches 71; Conservative 46; Mismatches 129; Indels 93; Gaps 14;
 QY 78 RDBPAKVVMETVEKMLADGVEYQFTFGGVPQGMIRVBGDTIEVQFSNHPDSK 137
 Db 9 RNYTFNVVMKNTT-----RLCS--SKPIVTVNGFPPPTLYABEDDVLVRVEN---RV 57
 QY 138 PHNVDFH-----AATGPGGAASFTAPGHTSTFSFKL--QPGLVVYCAVAVPG 186
 Db 58 KIVVTIHMGIROLRTMADGPAYITTCPI-QPQOSTVYVFTTIGCGTLLMHAHVI--- 112
 QY 187 MHIANWYGLILVEPKGL-----PKVDKEYVWQGFYTKGYG-----EOGLQPDME 236

Db 113 LMRATLHGAIVLLPRKGVYPPPKRKEVVVVLGEMMKSDTEGVIQAISKGLAD----- 168
 QY 237 KAIREDAEVVFNGSVGALT-----GENLAKKAGETVLLFVNGSPNLTSSHVIGEIF 221
 Db 169 ---NVSDAHTTNGHSPSSNCPSGGFTLPVESGKRWLRITNALNEELFFKIGH-- 222
 QY 292 DKVHFEKGGENHNIQTLLIPAGGAITEFKVDVPGDYVLVDHAIFFAPFKKALGILKYE 351
 Db 223 -----QITVEVDATYVKPKTDT-----IYIAP 246
 QY 352 GEENHEIYSHKQTDVYLLPEG-----APQALDTQAPKT 385
 Db 247 GQTTNALISTDQSGSKYVAASPFMDSPIAVDNMTATAT 285

RESULT 20
 US-09-252-991A-23061
 ; Sequence 23061, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 23061
 ; LENGTH: 320
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23061

Query Match 5.2%; Score 136; DB 4; Length 320;
 Best Local Similarity 25.7%; Pred. No. 0.00024;
 Matches 62; Conservative 25; Mismatches 80; Indels 74; Gaps 10;
 QY 195 GILVEPKGLPRVDKEYVWQGFYTKGKYEBOGLQPDMEKAIREDAEVVFNGSVGA 254
 Db 73 GYIALVPGIGTWKGLMGPVQSADDEFADKEX-GWTVHMQEKMA-KADEKYGIFPAKFAA 130
 QY 255 LTGENALK---AKVGETVRLVGN-----GPNLTSSFHVIGEIFDKVHFE 297
 Db 131 MPLEEVAKDPOAVKMG--RLFASNSICHSADAKAGVFPULTDPMRWG----- 179
 QY 298 GKGGENHNIQTLLIPAGGAITEFKVDVPGDYVLVDHAIFFAPFKKALGILKVEGENHE 357
 Db 180 ---GEBETIKTITM-AG-----RHAAMPWNGE---VIGEGVKNVA 213
 QY 358 IYSHKQTDVYLLPEGAPQALDTQAPKTPAPANLQEOIKAGKATYDSNCAACHQPDGKV 417
 Db 214 APLVTQMDGRKLPBGA-----KADIEAGQVFAITTCVACHGPEGKOT 255
 QY 418 P 418
 Db 256 P 256

RESULT 21
 US-09-032-315-8
 ; Sequence 8, Application US/09032315
 ; Patent No. 5985818
 ; GENERAL INFORMATION:
 ; APPLICANT: Svendsen, Allan
 ; TITLE OF INVENTION: LACCASE MUTANTS
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 59858180 No. 59858180disk of No. 59858180th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,315
FILING DATE: 27-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5200.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-032-315-8

Query Match 5.1%; Score 135; DB 2; Length 575;
Best Local Similarity 20.0%; Pred. No. 0.00077;

Matches 105; Conservative 63; Mismatches 176; Indels 180; Gaps 25;

QY 86 VKMEYVEKYMRLA-----DGEYQFWTFGGVPGQMIRREGDTIEVQFSN---HPDSK 136
DB 17 VLARTVEYGLKISDGEIAPDGVKRNATLVNGYFGLIPANKGDTLKVQNKLTNPEMY 76
QY 137 MHNVDVFNAA-----TGPGGGAASFT-----APGHTSTFSFPA-LQPGLYVYHCAY 182
DB 77 RTTSTHMGLOHRNADDGP-----SFVTCPIVPRRSYTYTLPDDQGTGYWH--- 127
QY 183 APVGNHIANGMVGLILVEPKE---GLPKYDKKEYVM--QGSFYTCKGYGQGLQPFMEKA 238
DB 128 SHLSQYVDGLGRLVYIPKDPHRLYVDDEKTVLIIGDWY-----HESSIA 175
QY 239 I-----REDAEYVFNQSVGALTGEN-----ALKAKVGETVRLFVNGGPNLTS 282
DB 176 ILASGNITRQRPVSATINCK-GRFPDNTNPANDTLTYTLKVRGKRYRLVINSSEIASF 234
QY 283 SFHVIGEIPDKVHFEKGKGNENIQTLLIPAGGAATTEKRVN----- 325
DB 235 RRSV-----EGHKV--TVIAADGVS TKPYQVDAFDILAGQRIDCVVEANQ 277
QY 326 -PGDYVL-----VDHAFRAFN---KGA 345
DB 278 EBDTYWINAPLTNVPKTAQALLVYEDRRPYHPKGPYRKSVSEALIKYNNHKKHGR 337
QY 346 GILKVEG-----EENHEIYS-----HKQTDAYVLPFGAPQALDIOEAPKTPAPANQ 392
DB 338 GILSGGGLKAMIEGSHLHSRYVKQNETTVVMDESKLVPLEYPGAAGSKPADLV 397
QY 339 EGIKAGKATYDNCACAPDQDGKGVNAPPL-----ANSYLANADHARAAS 439
DB 398 LDL-----TFGLNFAFGHMT-NGIPYESPKIPTLLKILITDEBGTESDFTKEH---TV 448
QY 440 IVANGISGKITVNGQYESVMPAIALSDQAIANVITYTLNFSFN 483
DB 449 ILPKKNCIEFNKNGSGIPITPHVHLGH-----TMDVVQGFEN 486

RESULT 22

US-08-993-318A-8
Sequence 8, Application US/08993318A
Patent No. 5998353

GENERAL INFORMATION:

APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Palle
APPLICANT: Rasmussen, Grethe
APPLICANT: Cherry, Joel
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 59983530 No. 59983530disk of No. 59983530th America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,318A
FILING DATE: December 18, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Gregg, Valera A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-318A-8

Query Match

Best Local Similarity 5.1%; Score 135; DB 2; Length 575;
Best Local Similarity 20.0%; Pred. No. 0.00077;

Matches 105; Conservative 63; Mismatches 176; Indels 180; Gaps 25;

QY 86 VKMEYVEKYMRLA-----DGEYQFWTFGGVPGQMIRREGDTIEVQFSN---HPDSK 136
DB 17 VLARTVEYGLKISDGEIAPDGVKRNATLVNGYFGLIPANKGDTLKVQNKLTNPEMY 76
QY 137 MHNVDVFNAA-----TGPGGGAASFT-----APGHTSTFSFPA-LQPGLYVYHCAY 182
DB 77 RTTSTHMGLOHRNADDGP-----SFVTCPIVPRRSYTYTLPDDQGTGYWH--- 127
QY 183 APVGNHIANGMVGLILVEPKE---GLPKYDKKEYVM--QGSFYTCKGYGQGLQPFMEKA 238
DB 128 SHLSQYVDGLGRLVYIPKDPHRLYVDDEKTVLIIGDWY-----HESSIA 175
QY 239 I-----REDAEYVFNQSVGALTGEN-----ALKAKVGETVRLFVNGGPNLTS 282
DB 176 ILASGNITRQRPVSATINCK-GRFPDNTNPANDTLTYTLKVRGKRYRLVINSSEIASF 234
QY 283 SFHVIGEIPDKVHFEKGKGNENIQTLLIPAGGAATTEKRVN----- 325
DB 235 RRSV-----EGHKV--TVIAADGVS TKPYQVDAFDILAGQRIDCVVEANQ 277
QY 326 -PGDYVL-----VDHAFRAFN---KGA 345
DB 278 EBDTYWINAPLTNVPKTAQALLVYEDRRPYHPKGPYRKSVSEALIKYNNHKKHGR 337
QY 346 GILKVEG-----EENHEIYS-----HKQTDAYVLPFGAPQALDIOEAPKTPAPANQ 392

Db 338 GLLSHGGLKARMBEGSHHLSRSVVRKONETTIVVNDSEKLVLEYPGAACGSKPADLV 397
Qy 393 EQIKAGKATYDSNCAACHQPDGKGVNAPFPL-----ANSDYLNADHARAAS 439
Db 398 LDL-----TGLNFAFGHMMI-NGIPYESPKEIPLTLKILTDEGVTESDFTKEEH---TV 448
Qy 440 IVANGLSKITVNGNOYESVMPALALSDQOIANVITYTLNSFCN 483
Db 449 ILPKNKIEFNKNGSGIPITHPVHLGH-----TWDVVOFCN 486

RESULT 23
US-09-399-886-8
Sequence 8, Application US/09399886
Patent No. 6140092
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Paule
APPLICANT: Raemussen, Grethe
APPLICANT: Cherry, Joel
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 61400920 No. 6140092disk of No. 6140092ch America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/399,886
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,318
FILING DATE: December 18, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-399-886-8

Query Match 5.1%; Score 135; DB 3; Length 575;
Best Local Similarity 20.0%; Pred. No. 0.00077;
Matches 105; Conservative 63; Mismatches 176; Indels 180; Gaps 25;

Qy 86 VKMETVKKWRLA-----DSGEVQFPTFGQVPGQMIKREGTIVQSN---HPDSK 136
Db 17 VLARTVEYGLKISGELTAPDGVKKNATLVNGVPGPLIPANKGTLVKVQNKLTNPEMY 76
Qy 137 MPNVNDEFAA-----TFPGGAEASFT-----AFGHTSFSPKA-LOGLYVYHCAV 182
Db 77 RTTSHHGHGLQHNADDDGP-----SFTVQCIVRRESTYTIPIPDQGTGYWYH--- 127
Qy 183 APYGMHIANGMVGLILVEPK---GLPKVDKEYVVM-QGDFYTKGKGEGLOPFDNEKA 238
Db 128 SHLSQVVDGLRGPLVIYKDPHRLRYDVDDKTVLLIGDMY-----HESSKA 175

Qy 239 I-----REDAEYVFNCSVGLTGEN-----ALKAKVGETVRLFVNGGPNLTS 282
Db 176 ILASGNITQRPVSATINCK-GRFPDNTPPADPTLYTLKVKRGKRYRLRVINSSSIAPF 234
Qy 283 SFHVIGEIFDKVHEGCKENHNIQTLIPAGGAITEKRVV----- 325
Db 235 RFSV-----EGHKV--TVIAAGSVSTKPYQVADFLLAQRIIDCVAEANQ 277
Qy 326 -PGDYVL-----VDHAFRANF---KCAL 345
Db 278 EPTTYWINAPLNVNPKTAQALLVYEDRRPYHPKPKYSVSEALIKYMHKHKGR 337
Qy 346 GILKVEG-----EENHEIYS-----HKQTDVAVLPEGAPQALDTQAPKTPAPANIQ 392
Db 338 GLLSHGGLKARMBEGSHHLSRSVVRKONETTIVVNDSEKLVLEYPGAACGSKPADLV 397
Qy 393 EQIKAGKATYDSNCAACHQPDGKGVNAPFPL-----ANSDYLNADHARAAS 439
Db 398 LDL-----TGLNFAFGHMMI-NGIPYESPKEIPLTLKILTDEGVTESDFTKEEH---TV 448
Qy 440 IVANGLSKITVNGNOYESVMPALALSDQOIANVITYTLNSFCN 483
Db 449 ILPKNKIEFNKNGSGIPITHPVHLGH-----TWDVVOFCN 486

RESULT 24
US-09-396-260-8
Sequence 8, Application US/09396260
Patent No. 6184015
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Xu, Peng
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 61840150 No. 6184015disk of No. 6184015ch America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/396,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/032,315
FILING DATE: 27-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5200,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-396-260-8

Query Match 5.1%; Score 135; DB 3; Length 575;
Best Local Similarity 20.0%; Pred. No. 0.00077;
Matches 105; Conservative 63; Mismatches 176; Indels 180; Gaps 25;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/172,331B
FILING DATE: 22-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/122,230
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/122,827
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/162,827
FILING DATE: 03-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lowney Dr., Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4052,020-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-172-331B-2

Query Match 5.1%; Score 133.5; DB 1; Length 576;
Best Local Similarity 20.2%; Pred. No. 0.0011;
Matches 106; Conservative 63; Mismatches 175; Indels 181; Gaps 26;
Db 86 VKRTEYKRLA-----DGEYQFTFGGVFGQVIRREGDTIVQSN--HPDSK 136
17 VLARTEYGLKISGEIAPDGVKNATLVAGYFGPLIFANKGDTLVKQNKLTNPEMY 76
137 MPNVDPHAA-----TSGGGAASFT---APGHTSTPSFKA-LDPGLVYHCAV 182
77 RTTSHHGLHNRNADDDG---SFTYQCTVRESYTYITPLDDGTGYWH--- 127
183 APVGMHANGMYG-LILVERKE--GLPKYDKEYVM--OGDPYTKGKGEGGLQPFDMK 237
128 SHLSQYVDLGRPLVYDPKDPHRLYVDDEKTVLIGDWY-----HSSK 175
238 AI-----REDAEYVFNQSVGALTGEN-----ALKAYGETVRLPVNGGPNLT 281
176 AIIASGNITRORPVSATINCK-GRFDPDNTNPANPDLTYTLKVKRKKVRLRVINSERIAS 234
282 SSFVIGEIFDKVHFEGSGENHNITLIPAGGAITEFKVDY----- 325
235 FRFSV-----EGHKV--TVIAADGVSTKPYQVDAFDILAQRIDCVVEAN 277
326 --PGDYVL-----VDHAFRAFN--KGA 344
278 QBPDTYINAPLTVNPKTAQALLVYEDRRPYHPRKGYRKMVSVAIIKKYNNHKHKG 337
345 LGILKVEG-----ENNHEIYS-----HKQTDAYTLREGARPAIDTOEAPKTPRANL 331
338 RGLISGCGKARMIESGSHLHRSVVKRONETTTVMDESKLPLEYPGAAGSKPADL 397
392 QEOIKKAKATYDSCAQCQDPDGVGNAPRPL-----ANSYLIADAHARA 438
398 VLDL-----TFGLAFATGHMMI-NGIPYESPKIPTLLIKILTDEGQTESDFTKEEH--T 448
439 SIYANGISGKITVNGQYESVMPAIALSDQOIANVITYTLNSFGN 483
449 VILPKKCIIEFNKNGSGIPITPHVHLGH-----TWDVVGQGN 487

RESULT 27

US-09-328-352-6003
Sequence 6003, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6003
LENGTH: 671
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6003

Query Match 5.0%; Score 132; DB 4; Length 671;
Best Local Similarity 20.4%; Pred. No. 0.0019;
Matches 96; Conservative 74; Mismatches 173; Indels 128; Gaps 23;
Db 107 TFCGVFGQVIRREGDTIEVQSNHPDSKMPHNVDFHATGPG-----GGAASFTAP 160
75 TVNGKFTAPLEFEBGDEAVIHVHNOQKQ-DTSLHMHGILLPGLMDGVPGFNNFKGIAP 133
161 GHTSPFRLALPGLVYHCAVAPVGMHANGMYGLILVERKGLP-----KVDKEYVM 215
134 NDFVYRPFVKQNGTYWH---AHSGQEGQDGLYGPVLPKGIIVAAHEKTRDYVM 190
216 QGDPYTKKKGEGGLQPFMEKAIREDAY-----VFNQSVGLTGE----- 258
191 LSDFHNSSS-----DSIMKRLKKAERYQNRRTVSVDLVKQVAGGLKATWDRSMW 242
259 ---NAKAKVGE-TVALFVNGS---PNTSSPHVIGEIFDKVHFGSGEHN-----N 305
243 NQMRMLKTDMSDVTGTYFLVNGTPOQNWTFNGFKADKQ--RIKFINASAMSEFVDIRN 300
306 IOTLLIPAGS-----AATFEKVDVPGDYVLVDHAFRAFNKAGLILVERGE--NHEIY 359
301 LKMTVVSADGQPVKPAIDEPRIETAETDYI-----VEPKQPVQI- 342
360 SHKQTDAYTLREGARPAIDTOEAPKTPRANLQEOIKKAKA--TYD-----SNCA 408
343 ---EASIS---DRSGAIGTLHNENTQAVPQMPILRRPALLTMDMGMSGGGSEHA 396
409 CHOPDGKVPNAFPPPL-----ANSDY-----LNAD-HARASIVANGSGKIT 450
397 GHQMNQHDMSAMPEKKETTOAHNDHTMKKDHDKMNSDGDHSHMNMHTHDSAMSE 456
451 VNGNQYESVMPAIALSDQOIANVITYTLNSFGNKGGLSADDAVAKKTKP 501
457 NMHD-----MQMMSASEHDHAMM-----NMNHEMPAQSENKPKQDEP 493

RESULT 28
US-09-328-352-8029
Sequence 8029, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8029
LENGTH: 635
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-8029

Query Match 5.0%; Score 131.5; DB 4; Length 635;

LENGTH: 540 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-687-5808-7

Query Match 4.9%; Score 129; DB 3; Length 540;

Best Local Similarity 22.3%; Pred. No. 0.0025;

Matches 100; Conservative 47; Mismatches 198; Indels 104; Gaps 20;

QY 100 GVEYQFTFGGQVGMIRREGDTIEVQSNHSDSPHNVDHATGCGGAEASF-- 157
 DB 42 GVPOQGLINGQFPGPDYISVTNDLINVNSLDE--PILLSWNGIQRNRSFEDGVG 99
 QY 158 ----TAGHTSTSEFKAL-QPGLVYHCAVAPVGMHANGWYG--LIVEPKEGLPYVD- 209
 DB 100 TTCEIPFGKRPFTYILQKQDQIGSYTFPSLA---FHAKAGFGKIKLSRPRIIPVPPDP 156
 QY 210 -KEYVMOGDPYTKGKYGEOGL-----QPFDEKAIREDAEVYVNGSVGALTGENALK 262
 DB 157 ADYTTLLIGDMYKKNHTALKALIDGCKLPF-----PDGILING-----RGPVGS 202
 QY 263 AKY--GETVALFVNGENPNTSSPHVIGEIPDKVHEFGKGEHNITTLIPAGGAITE 320
 DB 203 FVEQGGTYRLRISNVLQNSLNRIGHMKLVEVG---THTLQTYSSLDVHVGOT 258
 QY 321 FKVDVPGDYVLVDHAI-----FRAFNGALGILKEGENHEIYSHKQTDVAVYLPBGAP- 374
 DB 259 YSVLITADQAKHIVYVSSRFISQVLTITGVHL-----YSNNTVSGSGPPPGPT 309
 QY 375 -----QALDTEAPKTPAD-----ANLQEOIKA-----GRATYDNC 406
 DB 310 IQIDWSLNQARSIRITNSASGPRENPQGSYHYGMINTTRTIRLASAGQVNGKORYAVNS 369
 QY 407 AACHQPPGKQVNPAPPLANSIDYINADHABAASL---VANGLSGKITVNGQYESMPA 462
 DB 370 VSEVPILD-----TLKLIDYFKVGGFVSGISIDAPSGGGLFLDTSVLGADYQFIEI 421
 QY 463 IALSDQOIANYITYTLNSF-----GNKKG 486
 DB 422 VFENTEDI--VQSMHLNGYSFWVGMGCG 448

RESULT 31

US-09-252-991A-23088

; Sequence 23088; Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ. ID NOS: 33142

; SEQ. ID NO 23088

; LENGTH: 310

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23088

Query Match 4.8%; Score 125.5; DB 4; Length 310;

Best Local Similarity 22.0%; Pred. No. 0.0022;

Matches 68; Conservative 37; Mismatches 89; Indels 115; Gaps 15;

QY 187 MHANGWYGLIVPEKRGKLPKVDKEYVYVMOGDFTYTKKRYGEGQLQPFDEKAIREDAEVY 246
 DB 105 MNIAQEKYGPFA-----KYAMPLEEVAKDEHA-----MKQGSMPATYC 145

QY 247 -VFNGS--VQALTGENALAKAVGETVRLFVNGGPNLTSSFHYIGEIPDKVHEFGKGEN 303
 DB 146 SICHGSDAKAL-----GFPULADN-----EKRWG-GDP 173
 QY 304 HNIQTLLIPAGGAITEFKVDVPGDYVLVDHAIFFAFPKAGALGILKEGENHEIYSHKQ 363
 DB 174 QSIETITL--GGR-----HAIAMWG-----DILGEDVKNVAAVYRTE 210
 QY 364 TDVAVYPEGAPQALDTEAPKTPAPANLOEIKAGATYDSNCAACHOPDGK-----GVY 418
 DB 211 LAGLKIPGFT-----KADVAGAGQITSVNCVACHGEGKGTALVGAP 252
 QY 419 NAFPLANSIDYINADHABAASIVANGLSKITVNGQYESVMPAIA--LSDQOIANYITY 476
 DB 253 NLITNP--GAFYQSSVYAQLQOTIRHGRQO-----WPAQEPYIKKEKVHILAAV 259
 QY 477 TLNSFGNKG 485
 DB 300 IYNLSHNOG 308

RESULT 32

US-09-032-315-7

; Sequence 7; Application US/09032315

; Patent No. 5985818

; GENERAL INFORMATION:

; APPLICANT: Svendsen, Allan

; TITLE OF INVENTION: LACCASE MUTANTS

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 59858180 No. 5985818disk of No. 5985818th America, Inc.

STREET: 405 Lexington Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/032,315

FILING DATE: 27-FEB-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Rozek, Carol

REGISTRATION NUMBER: 36,993

REFERENCE/DOCKET NUMBER: 5200.200-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ. ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 572 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-032-315-7

Query Match 4.7%; Score 124; DB 2; Length 572;

Best Local Similarity 21.4%; Pred. No. 0.0081;

Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;

QY 86 VKAEYEVKRWRLA-----DGEYQFTFGGQVGMIRREGDTIEVQFSN---HPDSK 136
 DB 17 VLARTVEYNLKIISNGKIADPVERDATLVNGGYPGLPIANKGDTLKVQKQNTLTPDMY 76
 QY 137 MPANVDFAA-----TGGGGAASFTAGHTSTSEFK-ALQGLIVYHCAVAPVGM 187
 DB 77 RTTSIMHGLIOHRNADDDGPAFVVTOCPV-VPQASVYTMPADQGTGYWYH---SHLSS 132

QY 188 HIANGMYG-LILVERKEGLPKV-----DKEYYVMQGFY-TKKG----YBEOGLQPFDMK 237
DB 133 QYVDGLRGLVLYDPKDPHRLYDIDDEKTVLIGDMYHTSSKAILATNITLQCPDS-- 190
QY 238 AIREDAEYVVFNGSVGALNGEN-----ALKAVGETVRLFVNGGPNLTSSPHVIG 288
DB 191 -----ATINGK-GRFPDNTPANPNTLYTLKVRGKRYLRVINS--SALASF----- 235
QY 289 EIFDKVHEGKGENHNITOTLLIPAGGAITEFKVDVPGDYVLVDHAFRAFNKALGIL 348
DB 236 ----RMSIOGHK-----MTVIADGVSTKPYQVD----- 260
QY 349 KVEGEBNEIYSHKQTDVAYLPEGAPOAIDTOEAPKT---PAP-ANLQEOIKAGKATYDS 404
DB 261 -----SFDILAGORIDAV-----VEANOEPDTYWINAPLTNVANKTAQALLIYED 305
QY 405 NCAACHOPDG 414
DB 306 DRPYHPKPG 315

RESULT 33

US-08-993-318A-7
Sequence 7, Application US/08993318A
Patent No. 5998353
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Palle
APPLICANT: Rasmussen, Grethe
APPLICANT: Cherry, Joel
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59983530 No. 5998353disk of No. 5998353ch America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,318A
FILING DATE: December 18, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-318A-7

Query Match 4.7%; Score 124; DB 2; Length 572;

Best Local Similarity 21.4%; Pred. No. 0.0081;

Matches 79; Conservative 55; Mismatches 124; Indels 112; Gaps 20;

QY 86 VKMETYEKMRLLA-----DGEVYQFWTGGGVPGOMIRVREDTILEVQFSN---HPDSK 136
DB 17 VIARIVEYMLKISNGRIAPDVERDATLVNGGYVGBLIFANKDITLKVKVQNLTNPDMY 76

QY 137 MEHNVDFHAA-----TGPGGAASFTAGHTSTSEFK-ALQPGLYVYHCAVAPGM 187
DB 77 RTTSHMHGHLQHRNADDDGPAFVTCPI-VPQASVYTMPLGDGTGYWYH---SHLSS 132
QY 188 HIANGMYG-LILVERKEGLPKV-----DKEYYVMQGFY-TKKG----YBEOGLQPFDMK 237
DB 133 QYVDGLRGLVLYDPKDPHRLYDIDDEKTVLIGDMYHTSSKAILATNITLQCPDS-- 190
QY 238 AIREDAEYVVFNGSVGALNGEN-----ALKAVGETVRLFVNGGPNLTSSPHVIG 288
DB 191 -----ATINGK-GRFPDNTPANPNTLYTLKVRGKRYLRVINS--SALASF----- 235
QY 289 EIFDKVHEGKGENHNITOTLLIPAGGAITEFKVDVPGDYVLVDHAFRAFNKALGIL 348
DB 236 ----RMSIOGHK-----MTVIADGVSTKPYQVD----- 260
QY 349 KVEGEBNEIYSHKQTDVAYLPEGAPOAIDTOEAPKT---PAP-ANLQEOIKAGKATYDS 404
DB 261 -----SFDILAGORIDAV-----VEANOEPDTYWINAPLTNVANKTAQALLIYED 305
QY 405 NCAACHOPDG 414
DB 306 DRPYHPKPG 315

RESULT 34

US-09-399-886-7
Sequence 7, Application US/09399886
Patent No. 6140092
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Palle
APPLICANT: Rasmussen, Grethe
APPLICANT: Cherry, Joel
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 61400920 No. 6140092disk of No. 6140092ch America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/399,886
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,318
FILING DATE: December 18, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-399-886-7

Query Match	4.7%	Score 124	DB 3	Length 572
Best Local Similarity	21.4%	Pred. No. 0.0081		
Matches	79	Conservative	55	Mismatches 124, Indels 112, Gaps 20

QY	86	VKMETVEKMRILA-----	DGEYQFWPTFGGCVPGCMIRVREBDTIEVQFSN--	HPDSK	136
DB	17	VLARTVEYNLKISNCKIAPDVGERTATLVNGVYRPGPLIRANKGDTLKVQVQKRLNPDMY			76
QY	137	MPHNVDFPAA-----	TGPGCGEASTTAAHGHSSTFSFK-ALQRLYVNHCAVAFVGM		187
DB	77	RTTSHMHGGLLOHRNADDGPAFVPTQCP-IVQASVTYMPPLGDTGTGYWYH--	SHLSS		132
QY	188	HIANGMYG-LILVEPEGLRKY-----	DKEYYVMQSDFY-TKSK----	YGEQGLQPFDMEK	237
DB	133	QYVDLRLRPLIYDXKDPHRRLYXDIDDEKTVLIIQDWHYTSKALLATGNITLQOPDS--			190
QY	238	AIREDAEYVFNQSGALTGEN-----	ALKAKVETVLFPVNGGPNLTSPFHVIG		288
DB	191	-----ATINGK-GRPDPRMTPANPNTLYTLKVRGGRYRLRVINS--	SAIASF----		235
QY	289	EIFDKVHFEGGEGENHNTQTTLIRPGSAALTEPFKVDVDPDYLVLDHAIIRANRKAALGIL			348
DB	236	-----RMSIOGHK-----	MTVIAIDGVSTKRYQVD-----		260
QY	349	KVEGSENHETIYSHKOTDAVYLREGAFOALDIOEAPKT--	PAP-ANTLOEOLAKRATYDS		404
DB	261	-----SPDILAGQRIADV-----	VEANQEPDTYVNNPFLNVANKTAQALLIYED		305
QY	405	NCAACHQDPDG	414		
DB	306	DRRPHRPRKG	315		

RESULT 35
 US-09-396-260-7
 : Sequence 7, Application US/09396260
 : Patent No. 6184015
 : GENERAL INFORMATION:
 : APPLICANT: Svendsen, Allan
 : APPLICANT: Xu, Feng
 : TITLE OF INVENTION: LACCASE MUTANTS
 : NUMBER OF SEQUENCES: 10
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: No. 61840150 No. 6184015disk of No. 6184015th America, Inc.
 : STREET: 405 Lexington Avenue
 : CITY: New York
 : STATE: NY
 : COUNTRY: USA
 : ZIP: 10174
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: Fastseq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/396,260
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US/09/032,315
 : FILING DATE: 27-FEB-1998
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Rozek, Carol
 : REGISTRATION NUMBER: 36,993
 : REFERENCE/DOCKET NUMBER: 5200.200-US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 212-867-0123
 : TELEFAX: 212-878-9655
 : INFORMATION FOR SEQ ID NO: 7:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 572 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single

[illegible]

US-09-576-281-7
Sequence 7, Application US/09576281
Patent No. 6277611
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Palle
APPLICANT: Rasmussen, Grette
APPLICANT: Cherry, Joel
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6277611o No. 6277611disk of No. 6277611th America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/576,281
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,318
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123

NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 539 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-318A-1

Query Match 4.6%; Score 121.5; DB 2; Length 539;
Best Local Similarity 23.6%; Pred. No. 0.013;
Matches 92; Conservative 50; Mismatches 165; Indels 83; Gaps 23;

QY 112 VPGQMTIRREGDTIEVQFSN---HPDSKMPHNVDFHA-----ATGPGGAASFTAP 160
DB 49 VHGPIIRGGKNDNFELNVNVDLNPFTMLRPTSIHMGILFQGTNNWADGADVNCPI-SP 107
QY 161 GHTSTFSF-KALQPLGVYVHCAPVGMHIANGMVGLIIV-----EPKGLPKYDKEYVM 215
DB 108 GHAFVLYKFTPAHAGTGMWYH---SHFGTOYCDGLGPMVITYDNDPHALYDEDDENTII 164
QY 216 Q-GDFTTKKRYGEOGLQPFMEKAIREDAYVVFNGS---VGALTGE-NALKAKYGE--T 268
DB 165 TLADWVHIAPASIQG-----AAQPD--TLINGKRYVGGPAELSIIVNEOGKKYR 214
QY 269 VRLFVNGGPNLTSSFFHVGIFDKVHFEKGGENNIOTLLI PAGGAITEFKVDVPCD 328
DB 215 MRLISLSCDNNW--QPSIDHELTIIIVDQLTPEHTVDLQIFTGORYSFVLDANQP-- 270
QY 329 YVLVDHAIFRAF-NKGALGILKVEGENHEIYSHKQTDVYVLEPGAPQALDTEAPKTPA 387
DB 271 ---VDVWYIRAOQPNKGRNGLAGT-----FANGVNSALIRYAGANADPTTSA--NPN 317
QY 388 PAULOQIKAKGATYDSNCAACHOPDGKGVPNAPPLANSDYLNADHAPASIVANGSLG 447
DB 318 PAQLNE-----ADLHALIDPAAPGIFP--FGAAD--VNLRFQGFSS-----CG 356
QY 448 KITVNGNOYESVMPAI-----ALSDQOIAN 472
DB 357 RFTINGTAYES--PSVPTLLQIMSGAQSAN 384

RESULT 39
US-08-689-421-27
Sequence 27, Application US/08689421
Patent No. 6008029
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie S.
APPLICANT: Brown, Kimberley M.
APPLICANT: Kaupplinen, Sakari
APPLICANT: Halikier, Torben P
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING SAME
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: No. 60080290 No. 6008029disk of No. 6008029th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,421
FILING DATE: 9-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4554.204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 539 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-689-421-27

Query Match 4.6%; Score 121.5; DB 3; Length 539;
Best Local Similarity 23.6%; Pred. No. 0.013;
Matches 92; Conservative 50; Mismatches 165; Indels 83; Gaps 23;

QY 112 VPGQMTIRREGDTIEVQFSN---HPDSKMPHNVDFHA-----ATGPGGAASFTAP 160
DB 49 VHGPIIRGGKNDNFELNVNVDLNPFTMLRPTSIHMGILFQGTNNWADGADVNCPI-SP 107
QY 161 GHTSTFSF-KALQPLGVYVHCAPVGMHIANGMVGLIIV-----EPKGLPKYDKEYVM 215
DB 108 GHAFVLYKFTPAHAGTGMWYH---SHFGTOYCDGLGPMVITYDNDPHALYDEDDENTII 164
QY 216 Q-GDFTTKKRYGEOGLQPFMEKAIREDAYVVFNGS---VGALTGE-NALKAKYGE--T 268
DB 165 TLADWVHIAPASIQG-----AAQPD--TLINGKRYVGGPAELSIIVNEOGKKYR 214
QY 269 VRLFVNGGPNLTSSFFHVGIFDKVHFEKGGENNIOTLLI PAGGAITEFKVDVPCD 328
DB 215 MRLISLSCDNNW--QPSIDHELTIIIVDQLTPEHTVDLQIFTGORYSFVLDANQP-- 270
QY 329 YVLVDHAIFRAF-NKGALGILKVEGENHEIYSHKQTDVYVLEPGAPQALDTEAPKTPA 387
DB 271 ---VDVWYIRAOQPNKGRNGLAGT-----FANGVNSALIRYAGANADPTTSA--NPN 317
QY 388 PAULOQIKAKGATYDSNCAACHOPDGKGVPNAPPLANSDYLNADHAPASIVANGSLG 447
DB 318 PAQLNE-----ADLHALIDPAAPGIFP--FGAAD--VNLRFQGFSS-----CG 356
QY 448 KITVNGNOYESVMPAI-----ALSDQOIAN 472
DB 357 RFTINGTAYES--PSVPTLLQIMSGAQSAN 384

RESULT 40
US-09-399-886-1
Sequence 1, Application US/09399886
Patent No. 6140092
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
APPLICANT: Svendsen, Allan
APPLICANT: Schneider, Palle
APPLICANT: Rasmussen, Grethe
APPLICANT: Cherry, Joel
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: No. 61400920 No. 6140092disk of No. 6140092th America
STREET: 405 Lexington Avenue
CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/399,886
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/993,318
 FILING DATE: December 18, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Gregg, Valeta A.
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 5032.200-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 539 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-399-886-1

Query Match 4.6%; Score 121.5; DB 3; Length 539;

```

0Y      112  VPGQIIRVREGGTLIVQSN---HSDSKMHNVDFA-----ATGGGGAGAAFTPP 160
Db      49  VHGPIIRGGKINDFEELNVNDLDNPTLRLFTSIHHGLFQRTGNMADGADVNOCP1-Sp 107
QY      161  GHTSTFSF-KALQPLYYVHCAPARYGMIANGMTGLIV---EKEGELPKVDKEYYYM 215
Db      108  GHAFLYKFTLPAGHAGTFWYH---SHNGTGYCSGLRGPMVITYDDNDPHALYDEDDENTII 164
QY      216  Q-GDPTTGTGKIGEGLQRPDMEKALREDAEYVFNCS---VQALTGE-NALKAKGE--T 268
Db      165  TLAWMYHAPBSIQG-----AAQRPDA--TLINGKRGYVGGRPALELIVNVEQSKYR 214
QY      269  VRLFVNGSGPNLTSFHVIGEIDFKNHFEGKGENNIQTLLIPAGGAITEFKVDPGD 328
Db      215  MRLISLSCDPMN--OFSIDGHELTIIIVDGQLTEPTTVRLQITFCORYSFVLDAQP-- 270
QY      329  YVLVDHAIFFAAE-NKGALGILKEVEGENHEIYSHKQTDAYVLPREGARQAIIDTOEAKTPA 387
Db      271  ---VDNMYIRAQPNKGRNGLAGT-----FANGVNSAILRYAGAAANDPTTSA--NNN 317
QY      388  PAALQEQIKAGKATYDSNCACHQPDGKGVPAAPPLANSDYLANADHAPAASTIVANGLSG 447
Db      318  PAQLNE-----ADLHALIDPAAPIPIT--PGAAD---VNLRFQLGFS-----CG 356
QY      448  KITVNGNGYESVMPAI-----ALSQOQIAN 472
Db      357  RFTINGTAYES--PSVPTLLQIMSGAQSN 384

```

Search completed: August 27, 2003, 18:33:21
Job time : 33 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:19:41 ; Search time 84 Seconds
(without alignments)
948,580 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2628
Sequence: 1 MSKPLIKITTLICALSALML.....NKGQLSADDAKAKKTKEN 502

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq.13Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2626	100.0	502	22	AA67668
2	2626	100.0	502	22	AA67669
3	1388	52.9	392	24	ABP7627
4	1034	39.4	197	23	AA017561
5	349	13.3	924	22	AA742578
6	237	9.0	133	21	AA74274
7	237	9.0	184	24	ABP78130
8	222	8.5	132	21	AA74276
9	222	8.5	133	21	AA74275

10	172	6.5	1209	22	AA49440
11	163	6.2	478	14	AA33935
12	163	6.2	478	21	AA33598
13	163	6.2	579	12	AA14306
14	162	6.2	478	14	AA33288
15	150	5.7	374	24	ABJ18767
16	148	5.6	478	20	AA14051
17	146.5	5.6	555	12	AA12339
18	146	5.6	467	13	AA20193
19	146	5.6	559	16	AAW08380
20	145.5	5.5	468	12	AA13994
21	145	5.5	559	11	AA06518
22	144.5	5.5	581	23	AAU76443
23	141	5.4	529	16	AA172328
24	141	5.4	529	18	AAW16301
25	141	5.4	529	19	AAW76310
26	141	5.4	529	19	AAW60878
27	141	5.4	529	21	AAV50728
28	140.5	5.4	493	22	AAV93026
29	136.5	5.2	326	21	AA16365
30	136.5	5.2	557	19	AAW43448
31	136.5	5.2	576	16	AAW72318
32	135	5.1	575	19	AAW62501
33	134	5.1	575	19	AAW76313
34	132.5	5.0	541	23	ABP91862
35	131.5	5.0	511	22	AAW90799
36	131	5.0	497	22	AAW79105
37	130	5.0	554	21	AAW53312
38	130	5.0	554	23	ABP9295
39	130	5.0	2183	21	AAW03533
40	129.5	4.9	439	24	ABP78252
41	129.5	4.9	504	22	ABW00012
42	129	4.9	518	16	AAW76880
43	129	4.9	540	16	AAW76879
44	129	4.9	549	21	AAW49530
45	129	4.9	585	20	AAV05677

ALIGNMENTS

RESULT 1
ID AA67668
AA67668 standard; Protein; 502 AA.
AC AA67668;
DT 11-JUN-2001 (first entry);
DE Amino acid sequence of a BASB109 polypeptide.
KW BASB109; bacterial infection; vaccine; genetic immunisation.
OS Moraxella catarrhalis.
PN WO200119996-A1.
PD 22-MAR-2001.
PF 14-SEP-2000; 2000WO-EP09035.
PR 14-SEP-1999; 99GB-0021691.
PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
PI Thonnard J;
DR WPI; 2001-257883/26.
DR N-PSDB; AAF55658.
PT Novel BASB109 polypeptides of Moraxella catarrhalis useful for
diagnostic, prophylactic and therapeutic purposes against microbial
diseases, preferably bacterial infections

2-keto-D-glucuronate
Cytocrome C. Gu
Sorbitol dehydrog
Ascorbate oxidase
Sequence of cytoch
Pseudomonas aerugi
G. oxydans D-bioti
ADH complex protei
Brassica napus mic
A. altocetigenes me
Brassica microspor
Novel Botrytis cin
Laccase RSLac3. R
Rhizoctonia solani
Rhizoctonia solani
Rhizoctonia solani
R. solani laccase
C glutamicum prote
Eucalyptus grandis
Tobacco laccase c1
Laccase RSLact. R
Rhizoctonia solani
Rhizoctonia solani
Herbicidally activ
Corynebacterium gl
N. gonorrhoeae anti
Coprinus cinereus
Tobacco leaf anti
Arbidopsis thalia
Maize diphenyl oxi

XX Claim 3; Page 65; 93pp; English.

CC The present sequence represents a BASB109 polypeptide of Moraxella
CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for
CC treating bacterial infections, and as research reagents and materials
CC for the treatment of and diagnosis of diseases, particularly human
CC diseases. They are useful for inducing an immune response in an
CC individual, and to assess the binding of small molecule substrates and
CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,
CC and natural product mixtures. BASB109 polynucleotides are useful for
CC therapeutic or prophylactic purposes, in particular genetic immunisation
CC and in diagnosis of the stage and type of infection.

XX Sequence 502 AA;

Query Match 100.0%; Score 2626; DB 22; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.9e-227;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKPTLIKTTILICALSALMLSGCSNOADKAAQPKSSTVDAAKTANADNNAASQEHQGLP 60
DB 1 MSKPTLIKTTILICALSALMLSGCSNOADKAAQPKSSTVDAAKTANADNNAASQEHQGLP 60
QY 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFMTFGQVPGQMRVR 120
DB 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFMTFGQVPGQMRVR 120
QY 121 EGDTEIYQFSNHPDSKMPHNVDFHAATGPGGAEASFTAPGHTSTFSFKALQPLYYVHC 180
DB 121 EGDTEIYQFSNHPDSKMPHNVDFHAATGPGGAEASFTAPGHTSTFSFKALQPLYYVHC 180
QY 181 AVAPVGMHIANGMVGLIVEPKEGLPKVDKEYVMQGDFTYKKGYSQGLQPFDMKXIR 240
DB 181 AVAPVGMHIANGMVGLIVEPKEGLPKVDKEYVMQGDFTYKKGYSQGLQPFDMKXIR 240
QY 241 EDAEYVFNQSVGALITGENALAKKVGETVRLFVNGGPNLITSFHVIGEIFDKVHFEQK 300
DB 241 EDAEYVFNQSVGALITGENALAKKVGETVRLFVNGGPNLITSFHVIGEIFDKVHFEQK 300
QY 301 GENHNITOTLLIPAGGAITEFKVDVPGDYVLVDHAIKRAFNKALGILKVEGSENEHIYS 360
DB 301 GENHNITOTLLIPAGGAITEFKVDVPGDYVLVDHAIKRAFNKALGILKVEGSENEHIYS 360
QY 361 HKQTDVYLPBGAPQALDIOEAPKTPAPANLQEOIKAGKATYDSCAACHQPDGKVPNA 420
DB 361 HKQTDVYLPBGAPQALDIOEAPKTPAPANLQEOIKAGKATYDSCAACHQPDGKVPNA 420
QY 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQYESVMPALALSDQOIANVITYTLNS 480
DB 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQYESVMPALALSDQOIANVITYTLNS 480
QY 481 FGNKGQOLSADVDVAKAKTKPN 502
DB 481 FGNKGQOLSADVDVAKAKTKPN 502

RESULT 2
AAB67669
ID AAB67669 standard; Protein; 502 AA.

XX AAB67669;
XX 11-JUN-2001 (first entry)
DE Amino acid sequence of a BASB109 polypeptide.
XX BASB109; bacterial infection; vaccine; genetic immunisation.
XX Moraxella catarrhalis.
XX WO200119996-A1.

PD 22-MAR-2001

XX 14-SEP-2000; 2000MO-EP09035.

XX 14-SEP-1999; 99GB-0021691.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonard J;

DR WPI; 2001-257883/26.

DR N-PSDB; AAF55657.

PT Novel BASB109 polypeptides of Moraxella catarrhalis useful for

PT diagnostic, prophylactic and therapeutic purposes against microbial

PS diseases, preferably bacterial infections

XX Claim 3; Page 66; 93pp; English.

CC The present sequence represents a BASB109 polypeptide of Moraxella
CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for
CC treating bacterial infections, and as research reagents and materials
CC for the treatment of and diagnosis of diseases, particularly human
CC diseases. They are useful for inducing an immune response in an
CC individual, and to assess the binding of small molecule substrates and
CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,
CC and natural product mixtures. BASB109 polynucleotides are useful for
CC therapeutic or prophylactic purposes, in particular genetic immunisation
CC and in diagnosis of the stage and type of infection.

XX Sequence 502 AA;

Query Match 100.0%; Score 2626; DB 22; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.9e-227;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKPTLIKTTILICALSALMLSGCSNOADKAAQPKSSTVDAAKTANADNNAASQEHQGLP 60
DB 1 MSKPTLIKTTILICALSALMLSGCSNOADKAAQPKSSTVDAAKTANADNNAASQEHQGLP 60
QY 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFMTFGQVPGQMRVR 120
DB 61 VIDAIYTHAPEVPPVDRDHPAKVVMETVEKVMRLADGVEYQFMTFGQVPGQMRVR 120
QY 121 EGDTEIYQFSNHPDSKMPHNVDFHAATGPGGAEASFTAPGHTSTFSFKALQPLYYVHC 180
DB 121 EGDTEIYQFSNHPDSKMPHNVDFHAATGPGGAEASFTAPGHTSTFSFKALQPLYYVHC 180
QY 181 AVAPVGMHIANGMVGLIVEPKEGLPKVDKEYVMQGDFTYKKGYSQGLQPFDMKXIR 240
DB 181 AVAPVGMHIANGMVGLIVEPKEGLPKVDKEYVMQGDFTYKKGYSQGLQPFDMKXIR 240
QY 241 EDAEYVFNQSVGALITGENALAKKVGETVRLFVNGGPNLITSFHVIGEIFDKVHFEQK 300
DB 241 EDAEYVFNQSVGALITGENALAKKVGETVRLFVNGGPNLITSFHVIGEIFDKVHFEQK 300
QY 301 GENHNITOTLLIPAGGAITEFKVDVPGDYVLVDHAIKRAFNKALGILKVEGSENEHIYS 360
DB 301 GENHNITOTLLIPAGGAITEFKVDVPGDYVLVDHAIKRAFNKALGILKVEGSENEHIYS 360
QY 361 HKQTDVYLPBGAPQALDIOEAPKTPAPANLQEOIKAGKATYDSCAACHQPDGKVPNA 420
DB 361 HKQTDVYLPBGAPQALDIOEAPKTPAPANLQEOIKAGKATYDSCAACHQPDGKVPNA 420
QY 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQYESVMPALALSDQOIANVITYTLNS 480
DB 421 FPPPLANSDYLNADHARAASIVANGLSGKITVNGNQYESVMPALALSDQOIANVITYTLNS 480
QY 481 FGNKGQOLSADVDVAKAKTKPN 502
DB 481 FGNKGQOLSADVDVAKAKTKPN 502

SQ Sequence 197 AA;
 Query Match 39.4%; Score 1034; DB 23; Length 197;
 Best Local Similarity 100.0%; Pred. No. 1,2e-84;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKRTLLIKTLLICALSALMLSGCSNQADKAAQPKSSTVDAAKTANADNNAASQEHQELP 60
 1 MSKRTLLIKTLLICALSALMLSGCSNQADKAAQPKSSTVDAAKTANADNNAASQEHQELP 60
 DB 1 MSKRTLLIKTLLICALSALMLSGCSNQADKAAQPKSSTVDAAKTANADNNAASQEHQELP 60
 QY 61 VIDAIVTHAPBPVPVRDHPAKVVKETVEKMRDLADGVEYQFMTFGQVPGQMIRVR 120
 61 VIDAIVTHAPBPVPVRDHPAKVVKETVEKMRDLADGVEYQFMTFGQVPGQMIRVR 120
 DB 61 VIDAIVTHAPBPVPVRDHPAKVVKETVEKMRDLADGVEYQFMTFGQVPGQMIRVR 120
 QY 121 EGDTEVQFSNHPSPKPHNVDPHAATGPGGAEASTAPGHTSTFSFKALQPLGYVYHC 180
 121 EGDTEVQFSNHPSPKPHNVDPHAATGPGGAEASTAPGHTSTFSFKALQPLGYVYHC 180
 DB 121 EGDTEVQFSNHPSPKPHNVDPHAATGPGGAEASTAPGHTSTFSFKALQPLGYVYHC 180
 QY 181 AVAPVGMHIANGMVGLI 197
 181 AVAPVGMHIANGMVGLI 197
 DB 181 AVAPVGMHIANGMVGLI 197

RESULT 5
 AAU42578
 ID AAU42578 standard; Protein, 924 AA.
 AC AAU42578;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #3474.
 XX
 KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KM dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 PN WO200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-US12865.
 XX
 PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Sheiky YW, Persing DH, Mitcham JL, Wang SS, Bharia A;
 PI L'Alaisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR N-PSDB; AASS9518.
 DR
 DR WPI; 2001-616774/71.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 PS Claim 3; SEQ ID NO 3773; 1069pp; English.
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at fdp.wipo.int/pub/published_pct_sequences.

SQ Sequence 924 AA;
 Query Match 13.3%; Score 349; DB 22; Length 924;
 Best Local Similarity 31.4%; Pred. No. 5e-22;
 Matches 92; Conservative 46; Mismatches 121; Indels 34; Gaps 10;

QY 72 VPPPV--DRDHPAKVVKETVEKMRDLADGVEYQFMTFGQVPGQMIRVREGDTIEVQF 129
 650 VLPPLMTGRVHRWTLIAQ---ESVOEIAPETTIDAMTYNGRYMAPIHARIGDEMVRHL 705
 DB 72 VPPPV--DRDHPAKVVKETVEKMRDLADGVEYQFMTFGQVPGQMIRVREGDTIEVQF 129
 QY 130 SNHPDSKPHNVDPHAATGPGGAEASTF-----APGHTSTFSFKALQPLGYVYHCAYA 183
 706 VNR--GTMGSLDPFHAGT-----VSPTRVMKTTIAPGGLDYNFTLHRAIGLWLYHCSTA 756
 DB 184 PVGMHIANGMVGLIIVEPKKGLPKVDKEYVYVMQGFYTKGKYGEQGLQPFDMKAIRED 243
 757 PMSAHIAAGKFGAVIYPPHD-LPRADREFLVQSEYTLSEHNQAE---VNTKINETP 811
 QY 244 EYVFNQSVGALNGENALPAKVGETVRLFPVNGGPNLTSSFHVIGEIFDKVHEG----- 298
 812 DLTMFNGHANQVYFE-FLKARVERVRIWLAAGPFGSGSFHVVGTFQFVREGAVYTLK 870
 DB 299 -GKGENHNTQTTILPAGGAITEFKVDVPGDYLVVHAIFRANKALGILKY 350
 DB 871 RGNPEGGCGCALDLASAQGFVEMVEBERGRYFVNH5-FVEMERAKGFIEV 922

RESULT 6
 AAY74274
 ID AAY74274 standard; Protein, 133 AA.
 AC AAY74274;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria gonorrhoeae ORF 007 protein sequence SEQ ID NO:44.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides, AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 132 AA;

Query Match 8.5%; Score 222; DB 21; Length 132;

Best Local Similarity 42.7%; Pred. No. 7.3e-12;

Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

398 GKATYDSCAACHQPDGKGVNAPPLANSPLYLNADHARAASIVANGSGKITVNGNOYE 457

30 GQKVEESNCVACHGCKGEGRTMFPPLRYSDPFIMKKKQVLLSHWKGINGTIVNGKTYN 89

458 SWPATAISDQOIANVITTYLNSFGNKGQLSADVDYAKAKKTK 500

90 GFWPATASDADIAAATYIMNAPFDNGGSGVTEKDVQAKKSK 132

RESULT 9

AAV74275

AAV74275 standard; Protein; 133 AA.

AAV74275;

21-MAR-2000 (first entry)

Neisseria meningitidis ORF 007 protein sequence SEQ ID NO:46.

Neisseria meningitidis; *Neisseria gonorrhoeae*; antigen; vaccine;

antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

antibacterial; gene therapy.

Neisseria meningitidis.

MO957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.

31-JUL-1998; 98US-0094869.

02-SEP-1998; 98US-0098994.

02-SEP-1998; 98US-0099062.

09-OCT-1998; 98US-0103749.

09-OCT-1998; 98US-0103794.

09-OCT-1998; 98US-0103796.

25-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.

(GENO-) INST GENOMIC RES.

Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,

Peteren J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarelli M,

Tetzelin H, Venter JC;

WPI; 2000-062150/05.

N-PSDB; AA253037.

Novel *Neisseria* polypeptides predicted to be useful antigens for

vaccines and diagnostics

Claim 2; Page 180; 1453p; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides, AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 133 AA;

Query Match 8.5%; Score 222; DB 21; Length 133;

Best Local Similarity 42.7%; Pred. No. 7.4e-12;

Matches 44; Conservative 14; Mismatches 45; Indels 0; Gaps 0;

398 GKATYDSCAACHQPDGKGVNAPPLANSPLYLNADHARAASIVANGSGKITVNGNOYE 457

30 GQKVEESNCVACHGCKGEGRTMFPPLRYSDPFIMKKKQVLLSHWKGINGTIVNGKTYN 89

458 SWPATAISDQOIANVITTYLNSFGNKGQLSADVDYAKAKKTK 500

90 GFWPATASDADIAAATYIMNAPFDNGGSGVTEKDVQAKKSK 132

RESULT 10

AA48440

AA48440 standard; Protein; 1209 AA.

AA48440;

16-JUL-2002 (first entry)

2-keto-D-gluconate dehydrogenase.

Cell membrane bound; 2-keto-D-gluconate dehydrogenase; enzyme;

2,5-diketo-D-gluconate.

Unidentified.

Key Location/Qualifiers

Misc-difference 740 /note="Encoded by TG"

FT KR2000019366-A.

PD 06-APR-2000.

XX 10-SEP-1998; 98KR-0037413.

XX 10-SEP-1998; 98KR-0037413.

XX 10-SEP-1998; 98KR-0037413.

XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

XX Shin YC, Bahn JG, Yeom DY;

XX WPI; 2001-088033/10.

XX N-PSDB; ABL52917.

XX Base sequence of novel cell membrane-bound 2-keto-D-gluconate

XX dehydrogenase gene and method for conversion into

XX 2,5-diketo-D-gluconate - Noabstract

XX Claim 3; Page 9-11; 16pp; Korean.

XX The present invention relates to a novel cell membrane-bound

XX 2-keto-D-gluconate dehydrogenase and a method for conversion into

XX 2,5-diketo-D-gluconate. The present sequence is the protein sequence for

CC the 2-keto-D-gluconate dehydrogenase.

XX Sequence 1209 AA;

Query Match
Best Local Similarity 23.2%; Score 172; DB 22; Length 1209;
Matches 97; Conservative 53; Mismatches 153; Indels 116; Gaps 20;

154 EASFTAPGHTSTFS--FKALQPLVYVHCAVAP---VGMH-----IANGMYG---- 195
DB QACHTAPGATATFSGGVAIATSPMGVYISNITPSADGIFPGEAEFSQAVRHVRADGAG 839
QY 136 LILVEKGLPYVDKE-----YYWQDPFYTKKYGEGLO-PEPMERAI-----RE 241
DB 840 LTPAPAPYTSYSKITDEDLHALYVFMHGVKPVQOKRQTSLEPPFYLRFSEFMNMFAD 899
QY 242 DAEVY-----VFNGSV-----GALTGENALKAVGETVALLFVNGCP 278
DB 900 DPEYISDDSGSABMNKGNVLYVNGLAHCNTCHTPRGVLMQAGNRPAGAPIGSWY--AP 956
QY 279 NLTSSFFHVGIEIFDKVFEFGKGENHIQTTLIPAGAAITFEKVDPGVYLVDAIFR 338
DB 957 NITS-----DAISGIGWRNDELVO-----YIKTGRA--E 984
QY 339 AFNKALGILK-VEGEENHEIYSHKQTDVYLPPEGAP--QAIDTQEPAPKTPAPAMLOEQI 395
DB 985 GGNQAAGGMAEVAEHSLOYSLSDSLKAIVLYLSTTPIRDEGDTQPAVSFGKPADVENSI 1044
QY 336 KA-----GKATYDSNCAACHQPDGKGVPR-APPPLANSYLNADHARAASITYA 442
DB 1045 RGRNANNANHSLTNGALFSGNCASCHQPDGASANOAVPSLFHNT-ATGMENPAPLIAA 1103
QY 443 -----NGLSGKITVNGNGQYESVMPALISDQIANTVITYLNSFGNGKGLSADDA 494
DB 1104 ILFGVQRNTAAQGVLMPPG--FSSPSYVDKLSDAQVADISNFPVLAHNGNEVYVSAGDVA 1160

RESULT 11

AAR39355
ID AAR39355 standard; Protein; 478 AA.

XX AAR39355;

DT 13-JUN-1994 (first entry)

DE Cytochrome C.

KW C-553; improved production; prodn.; oxidative fermentation.

OS Gluconobacter suboxydans IFO 12528.

XX Key Location/Qualifiers

FT Peptide 1..36 /note= "signal peptide"

FT Peptide 37..316 /note= "mature peptide"

PN JP05049480-A.

PD 02-MAR-1993.

PF 14-JUN-1991; 91JP-0238579.

PR 14-JUN-1990; 90JP-0154096.

XX (ASAH) ASAH I CHEM IND CO LTD.

XX WPI; 1993-169634/21.

DR N-PSDB; AAC48234.

PT Cytochrome C gene - comprises specified sequence of 1473 bases
PT which is inserted in host cells and is used to produce cytochrome C
PT by oxidative fermentation

XX Claim 1; Page 13-15; 20pp; Japanese.

CC The sequence is that of encoded by the cytochrome C gene from
CC Gluconobacter suboxydans IFO 12528 which may be used to give
CC improved prodn. of cytochrome C by oxidative fermentation.

XX Sequence 478 AA;

Query Match
Best Local Similarity 36.0%; Score 163; DB 14; Length 478;

Matches 49; Conservative 18; Mismatches 59; Indels 10; Gaps 6;

QY 369 LPEGAPQAIIDTQEPAPKTPAPAMLOEQIYAGKATYDSNCAACHQPDGKGVNAPPLA-NS 427
DB 306 VEGKNLGG---QDDGQTALANKGGQGNAGAVYLHNCALICHMNDGTGNRPFPLAGNP 362
QY 428 DYLANDHAPASIVANGSGKITVNGNOYESV-MPAIA--LSDQIANTVITYTLNSFGNK 484
DB 363 VITDDPSTSLANVAVFG--GILPNTSAPSAVAMPGFKNHLSDOEMADVNFMRKGMGN 420
QY 485 G-GOLSADVDVAKAKKT 499
DB 421 APGTVASDIOKLRIT 436

RESULT 12

AAB35988
ID AAB35988 standard; Protein; 478 AA.

XX AAB35988;

DT 01-MAR-2001 (first entry)

DE Sorbitol dehydrogenase subunit 2 amino acid sequence.

KW Sorbitol dehydrogenase; SDH; Gluconobacter oxydans; D-sorbitol;

KW L-sorbose production; 2-keto-L-gulononic acid.

OS Gluconobacter oxydans.

PN WO200065066-A1.

PD 02-NOV-2000.

PF 23-APR-1999; 99WO-1B00736.

PR 23-APR-1999; 99WO-1B00736.

XX (CHOI/) CHOI E.

PA (RHEE/) RHEE S.

PA (LEE/) LEE E.

PI Choi E, Rhee S, Lee E;

DR WPI: 2000-687351/67.

DR N-PSDB; AAC83154.

PS Claim 6; Fig 8; 96pp; English.

CC This invention relates to an isolated membrane-bound sorbitol
CC dehydrogenase (SDH) from Gluconobacter oxydans. The invention includes
CC polynucleotides (AAC83153 - AAC83154) encoding the 3 subunits of SDH
CC (AAB35987 - AAB35989). Also included in the invention are two
CC polynucleotide sequences AAC83156 and AAC83157 which contain the subunit
CC coding regions. The sorbitol dehydrogenase polynucleotide sequences are
CC useful for producing L-sorbose from D-sorbitol and for increasing the
CC production of 2-keto-L-gulononic acid by transforming a host cell,
CC especially Gluconobacter with the DNA and selecting the transformed host

CC cell. Note: An additional 66 DNA sequences (fragments of the SDH subunit
CC genes) encoding fragments of SDH are specifically claimed, however these
CC sequences are not given separately in the specification but are included
CC in sequences AAC83156 and AAC83157.

SO Sequence 478 AA;

Query Match 6.2%; Score 163; DB 21; Length 478;

Best Local Similarity 37.3%; Pred. No. 1e-05; Mismatches 57; Indels 10; Gaps 6;

Matches 50; Conservative 17; Mismatches 57; Indels 10; Gaps 6;

QY 369 LPEGAPQAIPTQEPKTPAPANLQEQIKAGKATYDSNCAACHQPGKGVNAPPLANS 428

DB 306 VPEKKNIG---QDDGKATALLKAGCKDAGAEVYLHNCAICHMNDGTGNRMFPPLAGNP 362

QY 429 YLVNDHARA-ASTVANGSLGKITNGQYBSV--MPALA--LSPQIANVITYLTNSFGNK 484

DB 363 VLTIDNATISMANIVTFG--GILPPTNTPASVAMPFRDHLSDQIADVNVNFRKSWGNQ 420

QY 485 G-GQLSADVDVAKAK 497

DB 421 APGTLSSADIRKLR 434

RESULT 13

AAR14306 AAR14306 standard; Protein; 579 AA.

AC AAR14306;

DT 10-JAN-1992 (first entry)

DE Ascorbate oxidase.

KM Enzyme.

OS Cucurbita moschata.

PN JP03210182-A.

PD 13-SEP-1991.

PF 11-JAN-1990; 90JP-0005270.

PR 11-JAN-1990; 90JP-0005270.

PA (TOYM) TOYOBO KK.

DR WPI; 1991-314587/43.

DR N-PSDB; AAQ14178.

PT DNA encoding polypeptide with ascorbate oxidase activity - derived
PT from Cucurbita moschata, used to produce recombinant enzyme in high
PT yield.

PS Disclosure; Fig 1; 8pp; Japanese.

XX The sequence was deduced from a DNA sequence obtd. from a clone
XX isolated from a library prepd. from RNA extracted from the fruit
XX of C. moschata. The DNA can be used to produce the enzyme in high
XX yield.

XX Sequence 579 AA;

Query Match 6.2%; Score 163; DB 12; Length 579;

Best Local Similarity 24.6%; Pred. No. 1.4e-05;

Matches 83; Conservative 44; Mismatches 130; Indels 80; Gaps 19;

QY 101 VEXQFMT-----FCGQVPGOMTRVREGDITVEQFSV--HPDSKMPHNVDFHATG 148

DB 40 VETVFMAPDCNENIWMGINGQFPPTIRANAGDTVVELINKLTFEGVH--WHGILQ 96

QY 149 PG-----GGAASFTF--PGHTSTFSFKALQGLYVVCANVAPVGMHIANGMGLIVER 201

DB 97 RGRPMADGTASISQCAINCEFTFFVNFVDNPGTFEYH---GHGQORAGLGLSLVDP 153

QY 202 KEELPR---VDKEYVYMOQDFYTKKXGEQ-GIQPDMKAR--EDAEVYVNG----- 250

DB 154 PQGKKEPFHYDGEINILLSDWCHQSIHQEVGIS---SKPIRWIGEPQITLLNGGQDP 209

QY 251 -SVGALTGEN--ALKAKVGETVRLFVNGCGPNLTSSPHYIGEIFDKVHEGKGGENHNIQ 307

DB 210 CSTAAKYDSNLEBCKKAGSEPCAPYI-----FHMPPKTYAIRI-----AS 250

QY 308 TTLIPGGAITE--FKVDVPGDYV---LVDAHIFRANKGALGILVEGGEHEIY- 359

DB 251 TTLALNFAIGNHPLVVEADGNVQPFYTSIDIDYSESYSLITTDQNPSENVWSV 310

QY 360 ----SHKQTD----AVYLPEGAPQAIPTQEPKTPA 387

DB 311 GTRGRHPTPPGLTLLNLPNSYSK-LPTSPPETPA 346

RESULT 14

AAR32288 AAR32288 standard; Protein; 478 AA.

AC AAR32288;

DT 25-MAR-2003 (updated)

DT 08-JUN-1993 (first entry)

DE Sequence of cytochrome c-553 (CO).

KM Cytochrome; oxidative fermentation; electron transfer system.

OS Gluconobacter suboxydans IFO 12528.

PN Key Location/Qualifiers

FT Peptide 1..36

FT Region 37..478

PN FR2677665-A1.

PD 18-DEC-1992.

PF 12-JUN-1992; 92FR-0007131.

PR 14-JUN-1991; 91JP-0238579.

PA (ASAH) ASAH KASEI KOGYO KK.

DR Shimizu T, Takeda Y;

DR WPI; 1991-054531/07.

DR N-PSDB; AAQ36687.

PT Plasmid contg. gene for cytochrome C-553 (CO) - and transformed
PT cells useful in high productivity oxidative fermentation, e.g. of
PT ethanol to acetic acid

PS Disclosure; Fig 1; 39pp; French.

XX Purified cytochrome c-553 (CO) was isolation from Gluconobacter
XX suboxydans IFO 1258 and partially sequenced. From this information,
XX 32P-labelled oligonucleotide probes were constructed and used to
XX screen a bank of genomic EcoRI fragments from IFO 1258 in vector

XX 15p19RF. A plasmid which hybridised strongly contained an approx.
XX 1.5kb fragment which was sequenced. This sequence lacks the N-terminal
XX region which was located from the same source and used to construct

XX a complete gene. The probes used are based on the N-terminal amino
XX acid sequence (KGMGNNA) of peptide I which was isolated following

XX limited hydrolysis of c-553(CO) using arginyl endopeptidase.

XX (Updated on 25-MAR-2003 to correct PN field.)

SO Sequence 478 AA;

```

Query Match 5.2%; Score 163; DB 14; Length 478;
Best Local Similarity 36.0%; Pred. No. 1.2e-05;
Matches 49; Conservative 18; Mismatches 59; Indels 10; Gaps 6

OY 369 LPEGAPOAIDTQEAETPAANIAGQIKAKATYDSCNCAACHQDPGKGVPAFPPLA-NS 427
   |||
   :|||
Db 306 VPEGKNLG---QDDQGTALLNKGGQGNAGAVYLHNCAICHMNDGTGNRFPPLAGNP 362
   :|||
   :|||
OY 428 DYLNDAPRAASIVANGSGKTTVNGQYESV-MPALA--LSDQOIANYITTYTLNSFGNK 484
   :|||
   :|||
Db 363 VVIITDDPTMLGNVAVFG--GILPTNSAPSAAVAMPGFKNKLSDOEMADVNMFRKGMGNK 420
   :|||
   :|||
OY 485 G-GOLSAADVAKAKKT 499
   :|||
   :|||
Db 421 APTVTSASDIQKRTT 436

RESULT 15
ABJ18767
ID ABJ18767 standard; Protein; 374 AA.
AC ABJ18767;
XX
XX
DT 27-FEB-2003 (first entry)
XX
XX Pseudomonas aeruginosa biofilm formation-related protein #31.
DE
XX Biofilm formation modulation; biofilm-associated disease;
XX cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
XX catheter-associated infection; medical device-associated infection.
XX
XX Pseudomonas aeruginosa.
XX
XX WO200285295-A2.
XX
XX 31-OCT-2002.
XX
XX 19-APR-2002; 2002WO-US12532.
XX
XX 20-APR-2001; 2001US-285190P.
XX 24-OCT-2001; 2001US-344142P.
XX
XX (IOWA ) UNIV IOWA RES FOUND.
XX (HARD ) HARVARD COLLEGE.
XX
XX Whiteley M, Banger M, Lory S, Greenberg EP;
XX
XX WPI; 2003-075601/07.
XX N-PSDB; ABT14589.
XX
XX Identifying compound capable of modulating biofilm formation by
XX bacteria/bacterial antibiotic resistance, useful for treatment of
XX biofilm associated disease -
XX
XX Claim 1; Page 115-116; 154pp; English.
XX
XX The invention comprises a method for identifying a compound capable of
XX modulating biofilm formation by bacteria. The method of the invention is
XX useful for identifying a compound capable of modulating biofilm formation
XX by bacteria or modulating bacterial antibiotic resistance. The method of
XX the invention is also useful for diagnosing and treating a subject
XX (especially an immunocompromised human) that is afflicted with a biofilm-
XX associated disease or disorder, such as: cystic fibrosis; AIDS; middle
XX ear infections; acne; periodontal disease; catheter-associated
XX infections; and medical device-associated infections. The present amino
XX acid sequence represents a protein that is used in the invention.
XX
XX Sequence 374 AA;
XX
XX Query Match 5.7%; Score 150; DB 24; Length 374;
XX Best Local Similarity 22.0%; Pred. No. 0.0001;
XX Matches 63; Conservative 38; Mismatches 86; Indels 100; Gaps 11;

```

QY	265	VEGETARLVNGNGPRLTSSPHVIGIFPKVHEEGKGSENNHIQT---	TLIPAGGA-----	316
DB	135	LGQVVEYSNLTQF-----	---DQINRQAKDEHYLTLEVDPLVLPGTVKRFELI	182
QY	317	-----AITEF-----	KVDVPGDY-----	VLVDBA 335
DB	183	TSSDVIHSMWVPAFAVKRDAIPGFVNEAMTVVDEBGIIRGGCAELCGKHGMP	LTVD--	240
QY	336	IFRAFNKALGILKVEGSENNHIYSHKOTDAVYLEPAGAPQALDTQEA	PKTPAPANLQEOI	395
DB	241	-----VVKAKEFDQMLAKRKE-EAAKVE-----	LTSEKEMTK-----	EBLV 275
QY	396	KAGKATVDSNCAQHPDQKGVNPAFPLANSVDYLNADHARAASIVANGLSGKITV	NGQ	455
DB	276	ARGDKVYHTTCAACGQAEQGMPPMFALKSGSKVLTGPKENHLEVFNGVPG-	-----	327
QY	456	YESVMPAIA--LSDQILNVITTYTLNSFGNKGGLSADDAVAKAKTK	500	
DB	328	--TAMBAFGKQLENDLAAVITTERMANGNDGDMVTPKDVAAYAKQK	372	
RESULT 16				
ID	AAV14051			
ID	AAV14051	standard; Protein; 478 AA.		
XX	AAV14051;			
AC	AAV14051;			
XX	15-JUL-1999	(first entry)		
DT				
XX	G. oxydans D-sorbitol dehydrogenase.			
DE				
XX	D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulonic acid; precursor;			
KM	L-ascorbic acid production.			
XX				
XX	Gluconobacter oxydans.			
OS				
XX	W09920763-A1.			
PN				
XX	29-APR-1999.			
PD				
XX	13-OCT-1998;	98WO-JP04612.		
PF				
XX	17-OCT-1997;	97JP-0285280.		
PR				
XX	(FUJII) FUJISAWA PHARM CO LTD.			
PA				
XX	Ishii Y, Noguchi Y, Saico Y, Soeda S, Yoshikawa K;			
PI				
XX	WPI; 1999-302741/25.			
DR	N-PSDB; AAX57911.			
XX				
PT	Gene group for D-sorbitol dehydrogenase, useful for simple			
PT	large-scale production of L-sorbose or 2-keto-L-gulonic acid as			
PT	precursor for L-ascorbic acid			
XX				
PS	Claim 9; Page 57-59; 83p; Japanese.			
XX				
CC	This sequence represents the D-sorbitol dehydrogenase of the			
CC	invention. Cells transformed with a vector containing DNA encoding			
CC	the dehydrogenase can be used to produce L-sorbose or 2-keto-L-gulonic			
CC	acid as precursor for simple large-scale L-ascorbic acid production.			
XX				
XX	Sequence	478 AA;		
XX				
QY	Query Match	5.6%;	Score 148;	DB 20; Length 478;
	Best Local Similarity	21.9%;	Pred. No. 0.00023;	
	Matches 93;	Conservative 47;	Mismatches 146;	Indels 138; Gaps 20;
DB	180 CAVAPVGNHINMGVYGLILVPRKGL-----PKVDKEVYVWQGDFTYKGYGOGUQDPF	233		
	53 CHTPRGAGPRPG--GLVIASPMGIVASNTTPDPDGI-----GKYTEEFAN-	98		

QY 234 DMEKAIRED-----AEYVENGSVALTGENA-----LK 262
 DB 99 ALKGIARDGAMLYPAMPYTAYSIELADTDIALVYFMHGVAPLRQDNPKTELKFPENIR 158
 QY 263 AKYGETVRLVVG-----NGENLTSSF---HYIGEIPKVAHEGCKENNIQTTLL----- 311
 DB 159 AMMISNNLLPAGPPPAKGDPTQYSTERGHYLDAL-----GHGCTCHTPRNFLMGERS 212
 QY 312 -----PAGGAITEFKVDVGDYVLVDHAIFFRAFNKALG----- 346
 DB 213 SSAYLGCTPLAGWYAPNITTSNANSIGDSEDDLVQYLRGSGVGRQAAGAMNGEAVENS 272
 QY 347 ILKVEGENHEIYSHKQTDVAVYLPFGAPQAIIDTQ-----EAPKTPAP 388
 DB 273 FSLTDELDHAL-----AAVIRQ-IPKIEDSQAQPRDRFGVAVQPIVDLQKPKDRE 324
 QY 389 ANIOEQIKAGKATYDSNCAACHOPDGKGVNAPPEPLANSOYL-----NADHAPAASTIVANG 444
 DB 325 DDLFPM--DGERIYVNNCAACHGIDGAGAA DHFTPSLSNAVGAAPGADNLIIM--IVNG 380
 QY 445 LSGKITVNGNYESVMPAIA-----LSDQILANVITYTLNSFGNKGGLSADDDVAKAK 497
 DB 381 VDR--TTNG--HHVLMFGFPTSDVQRLSDTVAKLTVVSGTFGSGDHHVTAQDVAVAR 436
 QY 498 KTRP 501
 DB 437 EGGP 440

RESULT 17
 AAR12339
 ID AAR12339 standard; Protein; 555 AA.

XX AAR12339;
 DT 30-AUG-1991 (first entry)
 DE Acorbate oxidase from Cucumis sativus.
 KM Acorbate oxidase; transformant; polypeptide; activity.
 OS Cucumis sativus.
 PN JP03108485-A.
 PD 08-MAY-1991.
 XX 19-SEP-1989; 89JP-0244257.
 PF 19-SEP-1989; 89JP-0244257.
 PR 19-SEP-1989; 89JP-0244257.
 XX (TOYM) TOYOBO KK.
 PA (NAGS) NAGASE SANGYO KPL.
 DR WPI; 1991-180924/25.
 DR N-PSDB; AAQ12068.
 XX
 PT DNA encoding polypeptide having ascorbate oxidase activity -
 PT derived from Cucumis sativus, and transformants producing
 PT polypeptide in high yields
 PS Disclosure; Fig 2; 9pp; Japanese.
 CC The DNA is derived from Cucumis sativus and can be used to obtain
 CC large amts. of polypeptide having ascorbate activity.
 XX
 SQ Sequence 555 AA;

Query Match 5.6%; Score 146.5; DB 12; Length 555;
 Best Local Similarity 32.5%; Pred. No. 0.00039;
 Matches 40; Conservative 16; Mismatches 42; Indels 25; Gaps 6;
 QY 101 VEYQFWT-----FGGVPGQMTIRVBEGDITIEVQFSN--HPDSKMPHNVDFAATG 148

DB 13 VEYMFMSPCVENIVMGINGEFPGPTIRANAGDIVVVELTNKLHTEGVVH---WHGILO 69
 QY 149 PG-----GGAEASFTR--RGHTSPSPKALQPLLYHYHCAVAVPVGHHINMGWGLIVER 201
 DB 70 RGTPMADGTASISQCAINPGEITFYRFVVDKAGTYFYH---GHLGMQRSAGLGLSLVDP 126
 QY 202 KEG 204
 DB 127 PEG 129

RESULT 18
 AAR20193
 ID AAR20193 standard; Protein; 467 AA.

XX AAR20193;
 AC 16-APR-1992 (first entry)
 DT ADH complex protein (mol.wt. 44.000).
 DE Alcohol dehydrogenase; acetic acid; fermentation.
 KM Acetobacter alioacetigenes NH-24.
 OS Acetobacter alioacetigenes NH-24.
 PN JP03269988-A.
 XX 27-NOV-1991.
 PD 26-MAR-1990; 90JP-0073440.
 PF 26-FEB-1990; 90JP-0042301.
 PR 26-MAR-1990; 90JP-0073440.
 XX (NAKA-) NAKANO SUTEN KK.
 PA
 DR WPI; 1992-019325/03.
 DR N-PSDB; AAQ20384.
 XX
 PT Alcohol dehydrogenase complex structural gene - used in plasmid
 PT enhancing efficiency of acetic acid fermentation for
 PT transformed acetic acid bacteria
 PS Disclosure; Fig 4(1-3); 21pp; Japanese.
 CC Acetobacter transformed with the sequence encoding this protein can
 CC enhance the efficiency of acetic acid fermentation. The ADH complex
 CC can be easily extracted from the bacteria and purified and it can be
 CC used for the determination of an alcohol.
 CC See also AAQ20383-84, and -86-88.
 XX
 SQ Sequence 467 AA;

Query Match 5.6%; Score 146; DB 13; Length 467;
 Best Local Similarity 25.2%; Pred. No. 0.00033;
 Matches 84; Conservative 50; Mismatches 127; Indels 72; Gaps 19;

QY 184 PVGMHIANMGYGLI--VEPEKGLPKVDKEYVYVNGQDFYTKGYGEGQL-----OPFM-EK 237
 DB 149 PLSMRWPLGIWRMFPSPKDFTPAPGTDBEIRAGDYLVYGP--GHGAGCHTPPGFAMQEK 207
 QY 238 AIREDAEYVFNQSVGALTGEMALKAKYGETVR-----LFGVNGGPNLTSSFNHYIGIFDK 293
 DB 208 AL--DAA-----GGPDFLSGAPIDNWWAPSLNDPVGIGRWSEDDIYTFLLSGRI DHS 260
 QY 294 VHPFGGKGNHNIQTTLLPAGGAITEFKVDVGDYVLVD--HAIFRAFNKALGILKYVE 351
 DB 261 AVF--GGMGD-----VYAWSTQYFTDDDLHALAK--VLSLIPVPVPSQ 299
 QY 352 GEENHEIYSHKQTDVAVYLPFGAPQAIIDTQAPKTPAPANLQEQIKAKATYDSNCAACHQ 411
 DB 300 GN-----TYVDPSTANMLASG-----NTASVP-----GADTYVECAICHR 335

QY 412 PGGKVPNAPFPPLANSDYLNADHARA-ASIVANGLSGKITVNGQYBSV-MPAI--ALSD 467
DB 336 NNGGGVARMFPPLAGPVPVTENTPTSLVNVIAHG--GVLPPSNMWSAVAMPQYSLSLA 393
QY 468 QOIANVITYTLNFGNKG--GQLSADYAKAKKT 499
DB 394 QOIAVYNFIRTSWGNKAPGTVAADVTIKLRT 426

RESULT 19

AAW08380
ID AAW08380 standard; Protein; 559 AA.

XX AAW08380;

DT 07-APR-1997 (first entry)

DE Brassica napus microspore-specific clone L10 product.

XX Male sterile plant; hybrid seed; pollen; microspore; oilseed;

KW canola; rape; antisense RNA; stress resistance;

XX herbicide resistance; transgenic plant; crop protection.

OS Brassica napus.

PN WQ9640950-A1.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US08692.

PR 07-JUN-1995; 95US-0476864.

PA (PION-) PIONEER HI-BRED INT INC.

XX Albani DJ, Arnison PG, Fabijanski SF, Laurian RS;

DR WPI; 1997-052340/05.

XX N-PSDB; AAT48866.

XX Example 8; Fig 3b(1-4); 224pp; English.

CC A polypeptide sequence (AAW08380) was deduced from the coding

CC region contained in a fragment (AAT48863) of Brassica napus

CC microspore specific clone L10. Promoter regions of the L10 gene

CC have been used to develop pollen-specific vectors useful for

CC controlling antisense gene sequences for the prodn. of male

CC sterile plants and hybrid seed.

XX Sequence 559 AA;

Query Match 5.6%; Score 146; DB 18; Length 559;

Best Local Similarity 26.1%; Pred. No. 0.00043;

Matches 71; Conservative 37; Mismatches 132; Indels 32; Gaps 12;

QY 100 GVEYFWTGGGVPFGMIRREGDTIEVQFSNHPD-----SKMHNVDFFHAATGPGGG 152

DB 43 GVPQVILINGQFPFGHNINSTNNNVIYVFNLDLPFLITWNGIQRKNCWQDGTPTGM 102

QY 153 AEASTAPGHTSTFSKAL-OPGLVYHCAVAVNGHIANMGVGLVPERKGLP---K 207

DB 103 CP---IMPGTNYHFPQPDQIGSYFY---PTTGNHRAAGGAGGGRVNSRLIIPVYAD 156

QY 208 VDEKYYVMGDFYTKGKYEGLQPF-DMEKAI-REDAEVVFNQSVGALTGENA--LKA 263

DB 157 PEDDTYVLGLDWYTK---SHIQKRLDGRITGRPDG--IYVNGSKGDSADALFTL 211

QY 264 KYGETVRLFVNGGPNLTSSFHVIGEIFDKVHEGGKGNHNIQTTLIPAG--GAITE 320

DB 212 KPGKTVRIRICNVGVKTSINFRIOHNRKMLVENEGSHVLQNDYSLDVEVGCGFTIVTA 271
QY 321 FKVDVPGDYVLVDHAIFFRAFNKAGILKVEG 352
DB 272 NQ--EPKDYVMVASSRFLKTVITTTGLARYBG 301

RESULT 20

AAR13994
ID AAR13994 standard; Protein; 468 AA.

XX AAR13994;

DT 25-MAR-2003 (updated)

DT 09-DEC-1991 (first entry)

DE A.altoacetigenes membrane-bound ADH 44kD sub-unit.

XX alcohol dehydrogenase complex; carboxylic acid production;

KW cytochrome c; ss.

XX Acetobacter altoacetigenes.

PN EP448969-A.

PD 02-OCT-1991.

PF 26-FEB-1991; 91EP-0102793.

PR 26-MAR-1990; 90JP-0073440.

PR 26-FEB-1990; 90JP-0042391.

PA (NAKA-) NAKANO VINEGAR CO LTD.

XX (NAKA-) NAKANO VINEGAR CO LTD.

XX Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H;

PI Kawamura Y;

DR WPI; 1991-289462/40.

XX N-PSDB; AAQ13581.

XX This sequence was deduced from the nucleotide sequence isolated

CC from A.altoacetigenes MH-24 total DNA. The 44kD subunit of the

CC membrane-bound ADH complex is a cytochrome c.

CC See also AAQ13580 and AAQ13582-4.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 468 AA;

Query Match 5.5%; Score 145.5; DB 12; Length 468;

Best Local Similarity 36.4%; Pred. No. 0.00037;

Matches 39; Conservative 21; Mismatches 40; Indels 7; Gaps 5;

QY 398 GKATYDSNCAQCHQPDGKGVNAPFPPLANSDYLNADHARA-ASIVANGLSGKITVNGQY 456

DB 322 GADTVVKECALIHRNDGGGVARMFPPLAGNVPVTENTPTSLVNVIAHG--GVLPPSNMAP 379

QY 457 BSV-MPAI--ALSDQIANVITYTLNFGNKG--GQLSADYAKAKKT 499

DB 380 SAVAMPGYKSLISAQOIAVYNFIRTSWGNKAPGTVAADVTIKLRT 426

RESULT 21

AAR06518
ID AAR06518 standard; Protein; 559 AA.

XX AAR06518;

QY 233 --FDMEKAIRED A---EYVFN G S---VGALTGENA-----LKAKEVETVRLFWG 274
 DB 210 SVFEIWD SAROGAPAL EMTIANGTNIYDCSASTDANCVCCKFELTFVEGTRKRLRLI 269
 QY 275 NGC-----PMLTSSPHV-ICEITDKVHFECKEEN-----304
 DB 270 NVGIDSHFEFPALDNHTLVITANDLPVPIPTTDTLLIGQRYDVIVEANAAADNWMIRG 329
 QY 305 NIQT-----LIPAGAA-----ITEFKVDVPGDY 329
 DB 330 NMGITCSSSEANATGILRYDSSSTVDPTSVGTRGTCADBPVSLVPHLADV-GGY 388
 QY 330 VLVDAHIFRAF-----NKGALGI-----LKYEGEE-----NHEIYSHKQTD A-- 366
 DB 389 SLVDEQVSFAFTNYFTWTINSSSLIDWSSPTTLKIFNNETIPEPTDYNNVALNQTDAEE 448
 QY 367 --VYLP EG-----AQQALPTQAPKTPAPANI QEOIKAKKATYDSN 405
 DB 449 WVYVYLEDLTGFGIMWPHLHGHDFYVVAQETDVFSAITKSPANFNLPNRRDVALPGR 508
 QY 406 --CAACHOPDGKGVNAPFPPLANS DYLNADHARAASIVANGSGKITVNGNQVESVMPAI 463
 DB 509 GYLALAFKLDNPG-----SWLL--HCHIAWHA SGLAMQVESQSSI-----AI 550
 QY 464 ALSDOOI-----AMVITTYT 477
 DB 551 GMSDITDIFEDTCANMNAYT 569

RESULT 23

AAR72328
 ID AAR72328 standard; Protein; 529 AA.

XX AAR72328;

DT 25-MAR-2003 (updated)
 DT 14-OCT-1995 (first entry)

DE Laccase RSLac3.

KW RSLac3; laccase; lignin; lignosulfonate; polymerization;
 depolymerization.

OS Rhizoctonia solani.

PN W09507988-A1.

XX 23-MAR-1995.

PF 13-SEP-1994; 94WO-US10264.

PR 17-SEP-1993; 93US-0122230.

PR 17-SEP-1993; 93US-0122827.

PR 03-DEC-1993; 93US-0162827.

PR 22-DEC-1993; 93US-0172331.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PI (NOVO) NOVO-NORDISK AS.

PI Christensen BE, Schneider P, Wahlechner JA;

DR N-PSDB; AA086526.

XX WPI: 1995-131356/17.

XX New neutral Rhizoctonia laccase(s) and corresp. nucleic acids -

CC PT are used industrially for polymerising lignin, lignosulfate(s)

XX and phenolic cpds. and for oxidising dyes.

PS Claim 6; Page 44-46; 78pp; English.

CC sequenced to reveal a novel gene, RSLac3 (given in AA086526),
 CC encoding a laccase (AAR72328) optimally active at pH 6.0-8.5.
 CC Recombinant laccase was expressed in E. coli.
 CC (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 529 AA;

Query Match 5.4%; Score 141; DB 16; Length 529;
 Best Local Similarity 20.4%; Pred. No. 0.0011;
 Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPYIDAIYTHABEVPVPRDRHPAKVVK--METEKVRLADGEVYQWTFGGVPGQM 116
 DB 8 LPDLAAVST-----PAPAAYNNYKFDINNVAVPBGFORISVSVGLVPGTL 54
 QY 117 IYVRGEGDTLVQFSNNH---PDSKMPINVDH---AATPGGA EASFTAP-----GHNS 164
 DB 55 ITANKGDTLRINVTNOLDTPSMKRATTTHHGLFOATTDAEDGPAFTVTCPLAQLNLSYTY 114
 QY 165 TESFKALQGLVYVHCAPVGMHIANGMVY-G-LIIEP-----KEGLPKYDKEXYVMAQDPF 219
 DB 115 EIPLRG-QGTMMVH---AHLASQYVDGLRGLVIVDPNDPHKSRVDVDASTVWLEDM 170
 QY 220 Y-----TKGKYEEOG--LQPFMEKAIREDAEYVFN GSVGALTGENALKAKVGETV 269
 DB 171 YHTPAVLEKQMFSTNTTALLSPVDSGLINGKGRVY-----GSPAVPRSVINVKRGKRY 225
 QY 270 RLFVNGGPNLSSPFIIVEIDKVFHEGCKEENNIQTTLPAG-----AATPEKVD 324
 DB 226 RLRLVNASAI GSFYSI-----EGSHL--TVIEADGILHQPPLAVDSFQTL 268
 QY 325 VPGDYVLVDHAI FRAFNKALGILKVEGEBNHEIYSHKQTDVYLP EGAPQALDIOE--- 381
 DB 269 AGORISVIEANOTANVIRAPMTVAGAGTANADPTNVFALVHNEGAPNNEPTTEGGS 328
 QY 382 -----APKTPAPANI QEOIKAKKATYDS-----NCAACHOPDGKGV 417
 DB 329 AIGTALVEENLHALINPGAPGAPADVSLNALIGRSTVDGLIRFTFNNIKYEAP----- 383
 QY 418 PNAFPPLANS DYLNADHARAASIVANGSGKITVNGNQVESVMPALALSDOOIAMVITTYT 477
 DB 384 --SLPPL-----LKITLANNASNDADFTPEHTIVLP HNKVIELNITGGADHP 428
 QY 478 LNSFGNKGQGLSADVAKAKTKTPN 502
 DB 429 IHLRGH-----VPDIVSLGGTFN 447

RESULT 24

AAW16301
 ID AAW16301 standard; Protein; 529 AA.

XX AAW16301;

DT 07-AUG-1997 (first entry)

DE Rhizoctonia solani laccase isozyme 4.

KW Blue copper oxidase; laccase; enzyme engineering;

KW protein engineering; lignin depolymerisation; dye oxidation.

OS Rhizoctonia solani.

XX Key Location/Qualifiers

FT Binding-site 427 /note= "Type I copper site ligand"

FT Binding-site 470 /note= "possible Type I copper site ligand"

FT Binding-site 480 /note= "Type I copper site ligand"

FT Binding-site 485 /note= "Type I copper site ligand"

FT Peptide 217..226

A R. solani RS22 (IMI CC 358730) cDNA library was screened on

ABTS. The DNA insert of a plasmid from an isolated clone was

```

FT      /note= "preferred site for mutation (Claim 30),
FT      corresponds to W163 region of ascorbate
FT      oxidase"
FT      Peptide
FT      303..312
FT      /note= "preferred site for mutation (Claim 29),
FT      corresponds to R285 region of ascorbate
FT      oxidase"
FT      Peptide
FT      356..371
FT      /note= "preferred site for mutation (Claim 28),
FT      corresponds to Q353 and W362 region of
FT      ascorbate oxidase"
FT      Peptide
FT      416..429
FT      /note= "preferred site for mutation (Claim 26),
FT      contains Cu-ligating His"
FT      Peptide
FT      465..472
FT      /note= "preferred site for mutation (Claim 31)"
FT      Peptide
FT      466..470
FT      /note= "preferred site for mutation (Claim 18)"
FT      Peptide
FT      474..494
FT      /note= "preferred site for mutation (Claim 27),
FT      contains Cu-ligating His and Cys"
FT      Misc-difference
FT      379
FT      /note= "translated residue from gene sequence is
FT      Ile"
FT      Misc-difference
FT      415
FT      /note= "translated residue from gene sequence is
FT      Ile"
FT      XX
FT      MN09709431-A1.
FT      XX
FT      PD 13-MAR-1997.
FT      XX
FT      PF 03-SEP-1996; 96WO-US14087.
FT      XX
FT      PR 01-SEP-1995; 95US-0003142.
FT      XX
FT      PA (NOVO ) NOVO NORDISK BIOTECH INC.
FT      XX
FT      PI Berka RM, Wahleithner JA, Xu F, Berka R;
FT      XX
FT      DR WPI; 1997-192906/17.
FT      DR N-PSDB; AAT63317.
FT      XX
FT      PT New mutant blue copper oxidase enzymes - having different specific
FT      activities to wild-type enzymes, used for e.g. lignin
FT      (de)polymerisation or oxidation of dyes
FT      XX
FT      PS Disclosure; Fig 6A-D; 48pp; English.
FT      XX
FT      CC Rhizoctonia solani laccase isozyme 4 (AAM16301) can be engineered to
FT      alter e.g. its specific activity or pH-activity profile or to
FT      improve expression yields. The mutant laccase is prepd. by site-
FT      directed mutagenesis of the rsl4 gene (AAT63317) and expression in
FT      host cells. The mutation may be a deletion, insertion or pref. a
FT      substn. of one or more amino acids at a location no greater than
FT      12 Angstroms, pref. no more than 2.5 Angstroms, from the Type I
FT      copper site. The mutant enzyme can be used e.g. for the
FT      polymerisation or depolymerisation of lignins, oxidation of dyes,
FT      or polymerisation of phenolic or aniline cpds. present in liquids,
FT      e.g. for the treatment of fruit juices.
FT      CC
FT      XX
SQ      Sequence 529 AA;
Query Match 5.4%; Score 141; DB 18; Length 529;
Beech Local Similarity 20.4%; Fred. No. 0.0011;
Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;
QY      59 LPVIDAIYTHAPBEPVDRDHPAKVVK--METVEKWRLLADGVQFTFGGQVPGOM 116
DB      8 LPLAAVST-----PAPFAVRNKKPKDKVNVNAPDFQGRIVSVNGLVPGTL 54
QY      117 IRVREGDTIEVQFSNNH---PDSKMPHNVDFF---AATPGCGAEASFTAD-----GHTS 164

```

```

DB      55 ITANKDTRLRINTVNTQLTDPMSMRATTIHHGLFOATTADEGPAFVTCOPIAQNLSYTV 114
QY      165 TFSFKALQPLGVYVHCAPVPVGHNIANGMYG-LIIVEP-----KEGLPKYDKEXYVWQDPF 219
DB      115 EIRLGR-QIGSTMVYH---AHLASQYVDGLRGFLVITDPPDPKHSRYDVDDASTVWLEDM 170
QY      220 Y-----TKGKXGEGQ---LQPFMEKAIREDAAEYVNVGSGVLTGENALKAKVGETV 269
DB      171 YHTPAPVLEKQWSTNTNLTALSPVDSGLINGKRYV-----CGPAPRPSVIVNKKGRY 225
QY      270 RLFGVNGGPRLTSSPHVIGEIPDKVHFEGKGNNHIQTLTPAGQ-----AATPEKVD 324
DB      226 RLKVINASAIQSFTFSI-----EGHSL--TVLEADGILHQLPLAVDSFQY 268
QY      325 VPGDYVLVDHAIIFRAFNKGLGILKYEGBENHEIYSHKQTDVAYLPEGAPOAIDTQE--- 381
DB      269 AGQRYSVIYEANQTAANYVIRAPMTYAGACTNANLDPTVFVAVLHNEGAPNAEPTTEGQS 328
QY      382 -----AKTPAPANLQEQIKAKATYDS-----NCAACHQPDGKGV 417
DB      329 AIGTAVVEENVHALINPGAPGSGAPADVSLNLAIGRSTVDGILRFTFNNIKYEAP----- 383
QY      418 PNAFPPLANSDYLNADHARAASIVANGLSKITVNGNOYESVWPAALSDQOIANVITYT 477
DB      384 --SLPPL-----LKITLANNASNDADFTNEHTVLPNNIVIELNITGGADHP 428
QY      478 LNSFGNKGQLSADVDVAKAKTKKPN 502
DB      429 IHLHGH-----VFDIVKSLGCTPN 447

RESULT 25
AAM76310
ID AAM76310 standard; protein; 529 AA.
XX
XX AC AAM76310;
XX
XX DT 08-JAN-1999 (first entry)
XX
XX DE Rhizoctonia solani (I) laccase protein.
XX
XX KW Laccase; variant; oxidisation; dye transfer inhibition; bleaching;
XX KM denim; lignin modification; paper strengthening; phenol polymerisation;
XX KM hair dye; waste water treatment.
XX
XX OS Rhizoctonia solani.
XX
XX PN MN09838287-A1.
XX
XX PD 03-SEP-1998.
XX
XX PF 23-FEB-1998; 98WO-DK00070.
XX
XX PR 28-FEB-1997; 97DK-0000222.
XX
XX PA (NOVO ) NOVO-NORDISK AS.
XX
XX PI Svendsen A, Xu F;
XX
XX DR WPI; 1998-495393/42.
XX
XX PT New variants of Coprinus and related laccases with increased
XX PT oxidation potential - or altered pH optimum, or mediator or
XX PT oxygen-hydroxide ion pathways, useful for oxidation, for inhibiting
XX PT dye transfer and in bleaching textiles, especially as detergent
XX PT additive
XX
XX PS Disclosure; Pages 124-125; 147pp; English.
XX
XX CC The present sequence represents a laccase protein. The specification
XX CC describes active laccase variants (see AAM76282, AAM76296-99 and
XX CC AAM76316-17) having increased oxidation potential, altered pH optimum,
XX CC altered mediator and/or altered oxygen/hydroxide ion pathway.

```

CC laccase variants are used specifically to oxidise substrates, to
 CC inhibit dye transfer, and for bleaching textiles, specifically denim.
 CC They can also be used for lignin modification, strengthening paper,
 CC polymerisation of phenols, dyeing of hair and textiles and waste
 CC water treatment.

XX Sequence 529 AA;

Query Match 5.4%; Score 141; DB 19; Length 529;
 Best Local Similarity 20.4%; Pred. No. 0.0011;
 Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPEVPPVDDHDAKVVK--METVEKMRADGVYQFTFGQVPGQM 116
 DB 8 LPLAAVST-----PAPAAVRNYKFDIKVNVNAPDGFORSIVSNGLVGTL 54
 QY 117 IAVREDDTIEVOFSNH---PDSKMRNVDFH---AATGGGGAASFTAP-----GHS 164
 DB 55 ITANKGDTLRINVTQLTDPMSMRATTIHHHGLFOATTADEGPAFVTCPIAQNLSYTY 114
 QY 165 TEFKALQPLGVYHCAVAPVGMHIANMGY--LILVEP---KEGLPKVDKEYYVNGDPF 219
 DB 115 ELPRLRG-QGTGMKMYH---AHLASQYVDGLRGLVIYIDPNDPKSRVDVDASTVVMLEDM 170
 QY 220 Y-----TKGKYEQG---LQPFMEKAIREDAEYVVGSGVALTGENALKAVGETV 269
 DB 171 YHTPAVLEKQMFSTNNNTALISFPDPSGLINGKRVY---CGPAPRNVINVKGKRY 225
 QY 270 RLFVNGGNGLNLSFHVIGEIFPKAFEGGKGNHNIQTLLPAG---AATFERKD 314
 DB 226 RLKRVINASAIGSTFSI-----EGHSL--TVISADGILHQPLAVDSFOIY 268
 QY 325 VPDGVYLVVDHAIAPAFNKGALGILKVEGEENHEIYSHKQTDAYLLPEGAPQALDQGE-- 381
 DB 269 AGQRYVIVYEAQNTAANYVIRAMTVAGAGTANLDPNTVFAVLTGEGAPNAEPTTBQS 328
 QY 382 -----APKTPAPANTLQOIKAGATYDS-----NCAACHQPDGKGV 417
 DB 329 AIGTALVEENLHALINPAPGSGAPADVSLMLAIGRSTVGILRFENNIKYEAP----- 383
 QY 418 PNAFPPLANSDYIANADHARAASIVANGLSKITVNGQVSNVPAIALSSQOQIANVITTY 477
 DB 384 --SLPPL-----LKLILANNASNDADFTPEHETIVLPHNKVIELNITGGADHP 428
 QY 478 LNSFGKGGQLSADVDVAKAKTKTEN 502
 DB 429 IHLHGH-----VFIVKSLIGTIPN 447

RESULT 26
 AAM60878
 ID AAM60878 standard; Protein: 529 AA.

XX AAM60878;

XX 09-NOV-1998 (first entry)

XX Rhizoctonia solani (1) laccase.

XX Coprinus laccase-like enzyme; enzyme engineering; enzyme stability;
 KW detergent; bleaching.

XX Rhizoctonia solani.

XX Key Location/Qualifiers

FT Misc-difference 126 /note="variant may have Ala, Val, Leu, Ile, Pro,
 Phe, Met, Gly, Ser, Thr, Cys, Tyr, Asn,
 Gln, Asp, Glu, Lys, Arg or His (preferably
 Phe or His) at this position"

FT Misc-difference 135 /note="variant may have Ala, Val, Leu, Ile, Pro,
 Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,

FT

FT Gln, Asp, Glu, Lys, Arg or His (preferably
 Phe) at this position"

FT Misc-difference 127

FT /note="variant may have Ala, Val, Leu, Ile, Pro,
 Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,
 Gln, Asp, Glu, Lys, Arg or His (preferably
 Phe) at this position"

FT Misc-difference 171

FT /note="variant may have Ala, Val, Leu, Ile, Pro,
 Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,
 Gln, Asp, Glu, Lys, Arg or His (preferably
 Phe) at this position"

FT Misc-difference 76

FT /note="variant may have Ala, Val, Leu, Ile, Pro,
 Phe, Trp, Gly, Ser, Thr, Cys, Tyr, Asn,
 Gln, Asp, Glu, Lys, Arg or His (preferably
 Phe, Val, Ile, Leu or Gln) at this
 position"

XX MO9827198-A1.

XX 25-JUN-1998.

XX 16-DEC-1997; 97MO-DK00571.

XX 08-SEP-1997; 97DK-0001021.

XX 19-DEC-1996; 96DK-0001449.

XX (NOVO) NOVO-NORDISK AS.

XX Cherry JR, Pedersen AH, Rasmussen G, Schneider P;
 PI Svendsen A;

XX WPI; 1998-362768/31.

XX New laccase variants with improved stability - having amino acid
 PT changes based on Coprinus laccase structure, used for e.g.
 PT oxidation, dye transfer inhibition or bleaching

XX Claim 13; 147-148; 168pp; English.

XX This is a laccase enzyme of Rhizoctonia solani. The invention
 CC relates to the design of new variants of Coprinus-like laccases
 CC (see AAM60874-79, AAM60925 and AAM62501-03); the R. solani laccase
 CC shows 64.8% homology to C. cinereus laccase. The modifications
 CC are based on the previously unknown three-dimensional structure of
 CC C. cinereus laccase. Amino acid residues identified as being
 CC important to protein stability are identified and altered to
 CC improve stability. The variants are typically obtained by
 CC mutagenesis of laccase DNA and expression in a host cell. Variants
 CC are preferably modified within 15 (especially 10 or 5) Angstrom
 CC of a copper ion in the three-dimensional structure of the laccase.
 CC For R. solani laccases, preferred substitutions are
 CC one or more of W128F/H, Y137F, Y129F, Y171F or M78F/V/I/L/Q. The
 CC stabilised laccase variants can be used in detergent additives, for
 CC dye transfer inhibition in detergents, in bleaching of textiles (in
 CC particular denim), for lignin modification, paper strengthening,
 CC phenol polymerisation, hair dyeing and in waste water treatment.

XX Sequence 529 AA;

Query Match 5.4%; Score 141; DB 19; Length 529;
 Best Local Similarity 20.4%; Pred. No. 0.0011;
 Matches 103; Conservative 60; Mismatches 216; Indels 126; Gaps 20;

QY 59 LPVIDAIVTHAPEVPPVDDHDAKVVK--METVEKMRADGVYQFTFGQVPGQM 116
 DB 8 LPLAAVST-----PAPAAVRNYKFDIKVNVNAPDGFORSIVSNGLVGTL 54
 QY 117 IAVREDDTIEVOFSNH---PDSKMRNVDFH---AATGGGGAASFTAP-----GHS 164
 DB 55 ITANKGDTLRINVTQLTDPMSMRATTIHHHGLFOATTADEGPAFVTCPIAQNLSYTY 114

XX 18-DEC-2000; 2000EP-0127688.
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH68245.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 17; SEQ ID NO: 6780; 246bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: the sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 493 AA;
 Query Match 5.4%; Score 140.5; DB 22; Length 493;
 Best Local Similarity 19.0%; Pred. No. 0.0011;
 Matches 80; Conservative 57; Mismatches 174; Indels 109; Gaps 15;
 QY 15 LSALMLSGCSNQADKAAQPKSSTVDAAKTANADNAAQSEHOGELPVIDAIVTHAEVPP 74
 DB 12 LGLGLVLAG-----TGAVAACTSDPGPAA-----APG 38
 QY 75 PYDRDHPAKVYVMEVEKMKRL-----ADGVEYQPTFGGQVPGQMIRVREGDTIEVQ 128
 DB 39 PSLRPTPTALGEPTVRRLTLARPLSLDIGIEAKTWGVSDPTGDAIETAGDVLQVD 98
 QY 129 FSNH-PDSKMP--HNYDFHAATGPGGAASFTAPGHTSPFSKALQPLGVYVHCAPV 185
 DB 99 ITNELPESTSIHMHGALAHNAADGVGMOTDPLEPESBSYVEVPHGGTYFYH---SHT 155
 QY 186 GMIHANGMTGLILVEKEGLPKVDKEYYVMQGDFTYKKGXGEGLOPFDMEX----- 237
 DB 156 GQLDRGLHAPLIRPQDAEDQDVEMTYILDWVD---GIQGTDDDELKLTGWSGD 211
 QY 238 -----AIRDAEAYVFNVSVAL-----TGENALKAKVGETVRL-FV 273
 DB 212 HNGRMGMGHHGMMHGTTPRVLGGDVGVVPHYLLNGRIIPRAHRTFEARPGKARLRF 271
 QY 274 GNGGPNLTSFHYI--GEIFDKVHPEGKGENNIOTTLIPGAATFEKVDVP---GD 328
 DB 272 NSGGDTI---FKVALGCHRTVTHDGFVQPMETESIYSWGE-----RVDVEVILD 322
 QY 329 YVLVDHAIIFRAFNKALGILKVGEEENHE-----IYSHKOTDAVYIEGAP 374
 DB 323 GTFPLTALAAGKODRAFVIRTAGGAPRPDVDFPELSTGLLSLTKADRALLPBGRP 382

ID AAB16365 standard; Protein, 326 AA.
 AC AAB16365;
 XX
 XX
 DT 24-OCT-2000 (first entry)
 XX
 DE Eucalyptus grandis diphenol oxidase protein sequence SEQ ID NO:291.
 XX
 XX Plant; lignin biosynthetic pathway; Eucalyptus grandis;
 KM Pinus radiata; Monterey pine.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200022099-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 06-OCT-1999; 99WO-NZ00168.
 XX
 PR 09-OCT-1998; 98US-0169789.
 PR 14-JUL-1999; 99US-0143811.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Blokeberg LN, Havukkala IU;
 XX
 DR WPI; 2000-317962/27.
 XX
 PT Novel polynucleotide encoding enzymes involved in lignin-biosynthetic
 PT pathway useful for producing transgenic plants especially eucalyptus
 PT and pine species having altered lignin content, composition and
 PT structure
 PT
 PS Claim 18; Page 151-152; 213pp; English.
 XX
 CC The present invention describes isolated polynucleotides and proteins
 CC encoding and representing the enzymes cinnamate 4-hydroxylase (CAH),
 CC coumarate 3-hydroxylase (CHH), phenolase (PNL), O-methyl transferase
 CC (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase
 CC (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL),
 CC coniferol glucosyl transferase (CGT), coniferin beta-glucosidase (CBG),
 CC laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase,
 CC caffeic acid methyl transferase, caffeoyl CoA methyl transferase,
 CC coumarate CoA ligase, cytochrome P450 1XX1A, diphenol oxidase, flavanol
 CC glucosyl transferase, flavonoid hydroxylase, and isoflavone reductase,
 CC which are involved in the lignin biosynthetic pathway. The
 CC polynucleotides can be used for modulating lignin content, lignin
 CC composition and the structure of a plant, especially eucalyptus and pine
 CC species, and for modifying the activity of an enzyme involved in lignin
 CC biosynthetic pathway, and for producing a plant having altered lignin
 CC content, composition and structure. They can be used for designing probes
 CC and primers useful for detecting similar DNA and RNA sequences in any
 CC organism and for PCR amplification. The lignin content can be efficiently
 CC amplified using the polynucleotides. AAB67908 to AAB68201 and AAB16341 to
 CC AAB16449 represent polynucleotide and protein sequences used in the
 CC exemplification of the present invention.
 CC
 SQ Sequence 326 AA;
 Query Match 5.2%; Score 136.5; DB 21; Length 326;
 Best Local Similarity 20.9%; Pred. No. 0.0014;
 Matches 71; Conservative 46; Mismatches 129; Indels 93; Gaps 14;
 QY 78 RDHPAKVYVMEVEKMKRLADGVEYQFTFGGQVPGQIRVREGDTIEVFSNHDSTK 137
 DB 9 RNYTFVVMKNTT-----RLCS--SKPIYVANGMPGPTLYARBDTVLVRVSN---RV 57
 QY 138 PHNDPFH-----AATGGGAASFTAPGHTSTFSFAL--QPLGVYVHCAPV 186
 DB 58 KYNVTIHHGIRQLRTGWAADGPAVITQCP--QPGGSYYVNFITQRTGLNHAH---- 112
 QY 187 MHIANGMVGLILVEKEGL---PKVDKEYYVMQGDFTYKKGXG-----EGLOPFDM 236


```

Db      113 LMLRALHGAIVILPKRGVPEYFPKPKHEVVVVLGEWMSDTEGVISQAIKSLGAP---- 168
QY      237 KAIREDDEYVVFNSVQALT-----GENALKAKYGEIVRLPVGNGENLTSSPHVIGEIF 291
Db      169 ----NVSDAHTINGHPSPSSNCPSQGGFTLPVESGKYMRIINAALNEELFFKIAH-- 222
QY      292 DKVHFEKGKGENHNIQTLLI PAGGAITEPKVDVPGDYVLVDHAI PPAFNKALGILKVE 351
Db      223 -----QTIIVVDATYVKKPKFTDT-----IYIAP 246
QY      352 GEENHEIYSHKQTDAYVLPFG-----APQALIDQEAPKT 385
Db      247 GQTTNALISTDQSSGKYMAASPMDSPAIADNMNTATAT 285

RESULT 30
AAW43448 ID AAW43448 standard; Protein: 557 AA.
XX
XX AAW43448;
AC
XX
XX 01-JUN-1998 (first entry)
XX
XX Tobacco laccase clone pTL3 protein.
DE
XX Tobacco; laccase; oxidoreductase; probe; hybridisation; sycamore;
KM transgenic plant; digestibility; forage crop; paper manufacture;
KM lignin; parasite resistance.
XX
XX Nicotiana tabacum.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..22
FT /note= "signal peptide"
FT Protein 23..557
FT /note= "mature protein"

MO9745549-AI.
XX
XX 04-DEC-1997.
PD
XX
XX 30-MAY-1997; 97WO-FR00948.
PF
XX
XX 31-MAY-1996; 96FR-0006760.
PR
XX (CNRS ) CENT NAT RECH SCI.
PA
XX
XX Faye L, Gomord VM, Kiefer-Meyer MC, O'Connell A;
PI
XX WPI; 1998-032655/03.
DR N-PSDB; AAV01596.
XX
XX Modifying lignin biosynthesis in plants with gene encoding laccase
PT mRNA - or its anti-sense complement, especially for crops used as
PT fodder and for paper production
XX
XX Claim 19; Page 32-35; 72pp: French.
XX
XX This sequence represents a full length tobacco laccase (an oxidoreductase
CC acting on diphenols) encoded by the clone pTL-3. The complete protein
CC has a calculated molecular weight of 61.9 kD and a pI of 10.08. The
CC protein has a putative signal sequence of 22 amino acids, leaving a 535
CC amino acid protein of molecular weight 509.4 kD. The mature protein
CC contains 12 putative N-glycosylation sites (Asn-Xaa-Ser/Thr) and also
CC contains 2 potential copper ion binding sites. The gene sequence was
CC isolated from a cDNA library from tobacco leaves using a probe derived
CC from the sequence of the sycamore (Acer pseudoplatanus) laccase gene.
CC The gene can be used to transform plant cells for producing transgenic
CC plants having a reduced or altered lignin content, for improved
CC digestibility of forage crops, for paper manufacture. Plants with
CC increased lignin content may have better resistance to parasites.
XX

```

```

SQ Sequence 557 AA;
Query Match 5.2%; Score 136.5; DB 19; Length 557;
Best Local Similarity 20.5%; Pred. No. 0.0031;
Matches 89; Conservative 48; Mismatches 163; Indels 135; Gaps 17;

QY 107 TFGGQPGQMIRVREGDTIEVOFSNHPDSKMPHNVDF-----AATGGGAEAS 156
Db 47 TVNGKFPPTIYAREDDTVLVKVVNH-----VKNI.SIHMHGIRQLTGMADGPAYITQCP 102
QY 157 FTAPEGHTSFSPFAL-QPGLVYTHCAVAPVGMHIANGTGLIVERKEGL---PKVDXE 211
Db 103 I-QPGQNVYVNFITITQGRGTLFWHAIH-----IMLRATVGAIVILENLVPPFPKPNHE 157
QY 212 YVMQGDFTYTKGVRGEGQLQPFDMKAIRE-----PAEYVFNQSVG-----AL 255
Db 158 AVVILAEWMS-----DTEAVINEALKSLGAPVSDAHTI--NGHQPVSNCAS 204
QY 256 TGENALKAKYGEIVRLPVGNGENLTSSPHVIGEIPDKVHFEKGKGENHNIQTLLI PAGG 315
Db 205 QGGYKLVNDPCKTYMLRVYINAALNEELFFKIAH-----KMTVVEVDA 247
QY 316 AATFERKVDVPGDYVLVDHAI PPAFNKALGILKVEENHEIYSHKQTDAYVLPFG--- 372
Db 248 TYIKPKFTDT-----IYIAPGQTTNIVITANQSSGKYMAASPF 286
QY 373 --APQALIDQEAPKTPAPANLQEQIKAGKATYDSNCAHQPDGKGVPAFPPLANSDYL 430
Db 287 MDAPLAVDVTATATLHSGTGG-----NSHISLSTPCKATPAPANTFLDSRLS 337
QY 431 NA-----DHARAASI-----VANGLSGITVNGNQYESVMPALASDOOI 470
Db 338 NSKKYPAKVPKIKIDSLPFTVGLGINPCPTCKQNGSRVAVSNVNTFPMPTALLQAHF 397
QY 471 ANVITYTLNLSFGNGK 485
Db 398 -----FGTKG 402

RESULT 31
AAR72318 ID AAR72318 standard; Protein: 576 AA.
XX
XX AAR72318;
AC
XX
XX 25-MAR-2003 (updated)
DT 14-OCT-1995 (first entry)
XX
XX Laccase RSLacl.
DE
XX
XX RSLacl; laccase; lignin; lignosulfonate; polymerization;
KM depolymerization.
XX
XX Rhizoctonia solani.
OS
XX
XX MO9507988-AI.
PN
XX
XX 23-MAR-1995.
PD
XX
XX 13-SEP-1994; 94WO-US10264.
PF
XX
XX 17-SEP-1993; 93US-0122230.
PR 17-SEP-1993; 93US-0122827.
PR 03-DEC-1993; 93US-0162827.
PR 22-DEC-1993; 93US-0172331.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
PA
XX (NOVO ) NOVO-NORDISK AS.
PI
XX Christensen BE, Schneider P, Wahnlechner JA;
XX WPI; 1995-111356/17.
XX
XX N-PSDB; AAO86524.
DR

```

XX New neutral Rhizoctonia laccase(s) and corresp. nucleic acids -
PT are used industrially for polymerising lignin, lignosulphate(s)
PT and phenolic cpds. and for oxidising dyes.
XX
PS Claim 3; Page 30-33; 78pp; English.
XX
CC A R. solani RS22 (IMI CC 358730) genomic library was screened with
CC a probe based on consensus sequences of known laccases. The
CC RSLac1 gene given in AA086524 was isolated; it encoded a laccase
CC (AA082318) optimally active at pH 6.0-8.5. Recombinant laccase was
CC expressed in E. coli
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
CC Sequence 576 AA;
SQ
Query Match 5 2%; Score 136.5; DB 16; Length 576;
Best Local Similarity 20.4%; Pred. No. 0.0032; Index 181; Gaps 26;
Matches 107; Conservative 62; Mismatches 175; Indels 181; Gaps 26;
QY 86 VKMEYEVKWRLLA-----DGVEYQFWTFGGGVGMIRVREGDTIEVQFSN---HPDSK 136
DB 17 VLARTEYGLKISDGEIAPDGVGRNATLVNCGYGPGLIFANKDITLKVQNKLTNPENY 76
QY 137 MPRNVDFHAA-----TGGGGAASFT---APGHTSPFKA-IQPLIYVHCAY 182
DB 77 RTTSHMHGILQHRNADDGP-----SEVTQCPVPRSESYTYIPLDQGTGYVYH--- 127
QY 183 APVGMHIANMGV-LILVEPK---GLPKYDKYYVM-QGDFYTKRGKEGSLQPPDMK 237
DB 128 SHLSQYVDGLRGLVYDPKDPHRLYDVKDTVLIIIGDWY-----HESSK 175
QY 238 AI-----REDAYVVFNGSVGALTGEN-----ALKAYGETVLFVNGGPNLT 281
DB 176 ALASGNIRQRPVSATINGK-GRPPDMTPANPDLTYLTKYRGKRYLRVINSSEIAS 234
QY 282 SSFHYIGEIFDKYHFGKGENNIIQTLLIPAGAAITFEKDV----- 325
DB 235 PRFSV-----EGHKV--TVIADGVSTKPYOVADPILAGRIDVVEAN 277
QY 326 --PGDVVL-----VHAIFRAFN--KGA 344
DB 278 QEPDTYMINAPLTNVNKTQAQLLVYEEDRRPYHPKPKVSEALIKWNHKGKG 337
QY 345 LGILKEG-----EENHEIYS-----HKOTDAVYLPBGAPQALDQCAPKTPAPANL 391
DB 338 RGLLSGHGGLKARMTIGSHHSRSYVKQNTTYVMDESKLVPLEYGAACGSKRPADL 397
QY 392 QEOIKAGKATYDSCACHQPDGKGVNAPFPL-----ANSDYLNADHARA 438
DB 398 VLDEL-----TFGLNFATGHMMI-BGIPYSPKIPFLTKILITDEBDVTESDPTKEH---T 448
QY 439 SIYANGLSGKITVNGQYSEWPAIALSDQILANVTYTLNNGFN 483
DB 449 VILPNKKTIEFNKNGSGIPITHPVRLHGH-----TWVDVQGN 487
RESULT 32
AA062501
ID AA062501 standard; Protein; 575 AA.
XX
AC AA062501;
XX
DT 09-NOV-1998 (first entry)
XX
DE Rhizoctonia solani (TV) laccase.
XX
KW Coprinus laccase-like enzyme; enzyme engineering; enzyme stability;
XX detergent; bleaching.
OS Rhizoctonia solani.
XX
FH Key Location/Qualifiers

FT Misc-difference 411
FT /note= "variant may have Ala, Val, Leu, Ile, Pro,
FT Phe, Met, Gly, Ser, Thr, Cys, Tyr, Asn,
FT Gln, Asp, Glu, Lys, Arg or His (preferably
FT Phe or His) at this position"
FT
FT Misc-difference 125
FT /note= "variant may have Ala, Val, Leu, Ile, Pro,
FT Phe, Met, Gly, Ser, Thr, Cys, Tyr, Asn,
FT Gln, Asp, Glu, Lys, Arg or His (preferably
FT Phe) at this position"
FT
FT Misc-difference 134
FT /note= "variant may have Ala, Val, Leu, Ile, Pro,
FT Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,
FT Gln, Asp, Glu, Lys, Arg or His (preferably
FT Phe) at this position"
FT
FT Misc-difference 126
FT /note= "variant may have Ala, Val, Leu, Ile, Pro,
FT Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,
FT Gln, Asp, Glu, Lys, Arg or His (preferably
FT Phe) at this position"
FT
FT Misc-difference 170
FT /note= "variant may have Ala, Val, Leu, Ile, Pro,
FT Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,
FT Gln, Asp, Glu, Lys, Arg or His (preferably
FT Phe) at this position"
FT
FT Misc-difference 75
FT /note= "variant may have Ala, Val, Leu, Ile, Pro,
FT Phe, Trp, Gly, Ser, Thr, Cys, Tyr, Asn,
FT Gln, Asp, Glu, Lys, Arg or His (preferably
FT Phe, Val, Ile, Leu or Gln) at this
FT position"
FT
FT MO0827198-A1.
FT
FT 25-JUN-1998.
FT
FT 16-DEC-1997; 97WO-DK00571.
FT
FT 08-SEP-1997; 97DK-0001021.
FT
FT 19-DEC-1996; 96DK-0001449.
FT
FT (NOVO) NOVO-NORDISK AS.
FT
FT Cherry JR, Pedersen AH, Rasmussen G, Schneider P;
FT Svendsen A;
FT WPI; 1998-362768/31.
FT
FT New laccase variants with improved stability - having amino acid
FT changes based on Coprinus laccase structure, used for e.g.
FT oxidation, dye transfer inhibition or bleaching
FT
FT Claim 16; 152-154; 168pp; English.
FT
FT This is a laccase enzyme of Rhizoctonia solani. The invention
FT relates to the design of new variants of Coprinus-like laccases
FT (see AA060874-79, AA060925 and AA062501-03); this R. solani laccase
FT shows 59.7% homology to C. cinereus laccase. The modifications
FT are based on the previously unknown three-dimensional structure of
FT C. cinereus laccase. Amino acid residues identified as being
FT important to protein stability are identified and altered to
FT improve stability. The variants are typically obtained by
FT mutagenesis of laccase DNA and expression in a host cell. Variants
FT are preferably modified within 15 (especially 10 or 5) Angstrom
FT of a copper ion in the three-dimensional structure of the laccase.
FT For R. solani laccase (II) variants, preferred substitutions are
FT one or more of W11F/H, Y125F, Y134F, Y126F, Y170F or M78F/V/I/L/Q.
FT The stabilised laccase variants can be used in detergent additives,
FT for dye transfer inhibition in detergents, in bleaching of textiles
FT (in particular denim), for lignin modification, paper strengthening,
FT phenol polymerisation, hair dyeing and in waste water treatment.
FT
FT Sequence 575 AA;

Query Match 5.1%; Score 135; DB 19; Length 575;
 Best Local Similarity 20.0%; Pred. No. 0.0044;
 Matches 105; Conservative 63; Mismatches 176; Indels 180; Gaps 25;

Oy	86	VGMETVEYKXMRLLA-----	DGVYQWMTREGQVPGQMIRREDDTLEVGNS--	--HPDSK 136
Dd	17	VLARIVEYGLKISDGEIAPDGVRRNATLVNGSYFGPLIFANKQDTLKVQNKLTINPEMY	76	
Oy	137	MEHNVDFHAA-----	TGPGGAASFT----	APGHTSFSPKA-LQPGLYVYHCAV 182
Dd	77	RTTSHMHGLLQHRNADDGP-----	SFVTQCPIVRESRYTYTIPLDDQTGYWYH----	127
Oy	163	APVGHIAIANGMYGLLIVEPKE---	GLPKYDKXYVW--QGFPTKTKKXGEGGLQPFOMEKA	238
Dd	128	SHLSQYVYGLKGLPLVITPDKPHRLYDVDDKTVTLIGDWY-----	HESSKA 175	
Oy	239	I-----	REDEYVVFNGSVGALITGEN-----	ALAKAYGETVRLFPVNGGPNLTS 282
Dd	176	ILASGNITQRVPASATINCK--	GRFPDDNPANPDLYTLKVKRKGKRYRLVINSSEIASF	234
Oy	263	SHVYIGEITDKVHFEKGKGNHNIQTLLPAGAAITFEKVDY-----	325S	
Dd	235	RSV-----	EGHKV--TVIADVSTKPYVDADFILAGRIDCVVEANQ	277
Oy	326	-GQDYVL-----	VDHAFAPFN-----	KGAL 345
Dd	278	EDDTYINAPLPTNVPRKTAQALLVEEDRRPHYPRKGYRKMSVSEALIKYNNHKKHGR	337	
Oy	346	GILKVEG-----	EENHEIYS-----	HKQDAVYLLPEGAPQALDTPQAPKTPAPANLQ 392
Dd	338	GLSGHGGLKAMWIEGSHLHSRSVVKRONETITVWDESKLVPLRYPGAACGSKPADLV	397	
Oy	393	EDIKAKATYDSCAICHQPDGKGVNAPPL-----	ANSDYLNADHARAAS	439
Dd	398	LDL-----	TFGNLFATGHMMI--NGIYESKFIPTLLKILTDGQVTSDDTKEEH--	TV 448
Oy	440	IVANGLSGKITVNGNGYESVMPAIALSDQIANVITYTLNLSFGN	483	
Dd	449	ILPKKICIEFNIGKNSGIPITHPVHLGH-----	TWDVQVQFGN	486
RESULT 33				
AAW76313				
ID	AAW76313	standard; proteain; 575 AA.		
XX	AAW76313;			
AC	AAW76313;			
XX				
DT	08-JAN-1999	(first entry)		
XX				
DE	Rhizoctonia solani (IV)	laccase proteain.		
XX				
KW	laccase; variant; oxidation; dye transfer inhibition; bleaching;			
KW	denim; lignin modification; paper strengthening; phenol polymerisation;			
KW	hair dye; waste water treatment.			
XX				
OS	Rhizoctonia solani.			
XX				
PN	WO983287-A1.			
XX				
PD	03-SEP-1998.			
XX				
PF	23-FEB-1998;	98WO-DK00070.		
XX				
FR	28-FEB-1997;	97DK-0000222.		
XX				
PA	(NOVO) NOVO-NORDISK AS.			
XX				
PI	Svendesen A, Xu F;			
XX				
DR	WPI; 1998-495393/42.			
XX				
PT	New variants of Coprinus and related laccases with increased			

PR oxidation potential - or altered pH optimum, or mediator or
PR oxygen-hydroxide ion pathways, useful for oxidation, for inhibiting
PR dye transfer and in bleaching textiles, especially as detergent
PR additive

PS Disclosure; Pages 129-131; 147pp; English

CC The present sequence represents a laccase protein. The specification
CC describes active laccase variants (see [AAW76282](#), [AAW76296-99](#) and
CC [AAW76316-17](#)) having increased oxidation potential, altered pH optimum
CC altered mediator and/or altered oxygen/hydroxide ion pathway. The
CC laccase variants are used specifically to oxidise substrates, for
CC insoluble dye transfer, and for bleaching textiles, specifically denim.
CC They can also be used for lignin modification, strengthening paper,
CC polymerisation of phenols, dyeing of hair and textiles and waste
CC water treatment.

SQ Sequence 575 AA;

Query Match	5.1%	Score 134	DB 19	Length 575
Best Local Similarity	20.0%	Pred. No. 0.0054		
Matches 105	Conservative 63	Mismatches 176	Indels 180	Gaps 25

[illegible]

XX WO200210210-A2.
PN
XX 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI; 2002-269010/31.
XX
XX
XX Identifying plant target proteins for herbicidally active compounds,
PT comparing aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
XX Claim 5; SEQ ID NO 1073; 261bp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
XX
SQ Sequence 541 AA;

Query Match 5.0%; Score 132.5; DB 23; Length 541;
Best Local Similarity 22.4%; Pred. No. 0.0066;
Matches 88; Conservative 49; Mismatches 173; Indels 83; Gaps 18;

QY 100 GVEYQWTFGGQVPGQMIRVREGDTIEVQSNHPS-----KMFNVDPHAAT 147
DB 45 GVPQVILINGQFPFPIEGCVTNNIVVINKLDEPILTNIGIKQKMSMODVLTGN 104
QY 148 GGGGAASFTAPGHTSTFSFKAL-OPGLVYHCAVAPGMHANGYGLIIVEPEGL- 205
DB 105 CP-----IQKSSWTHFQKQDQIGTAY--FASISMRAASAFGLANNQGSVIF 153
QY 206 ---PKYDKYVYMGDFYTKGKYGEGLO-PFMKAIREDAYVFNQSVG-ALTGENA 260
DB 154 VEPKRDADFTLLVSDMY--KMGHKLQRLDSSRALPPDGLLNGASKGLVFTGQ- 209
QY 261 LKAXGETRLFPVNGGPNLTSFHHVIGELFIDVHFEKGKGENHNIOTL---IPAGA 316
DB 210 ---GKIRFRISNIGISTINFRIGHMTLVEEG---SHTIQEYESLIDHVGOS 260
QY 317 ATTEFKVDPV-GDYVLVDHAIFFAFKAGLGLIKVEEENHEIYSHKQTDVAVLPFGAQ 375
DB 261 VTVLVTLKAPVDYFIVASTRTFKPLITTTGILSYGSKRP--SHP-----LPIGPY 312
QY 376 ALDTEAPETPAPANIQEOIKAGKATYDNCACHPDGGKGVNAPPLANSDYLNADHA 435
DB 313 HHWS-----MKQARTIRLMLTANAA--RPNPGS-----SFHYGTIP 348
QY 436 RAASIYANG---LSKITVNGNGYESVMPAIL 465
DB 349 NRTFVLNBRAMINGKLRITYVNVSVYVNPATPL 381

RESULT 35
ID AAG90799
XX AAG90799 standard; Protein; 511 AA.
XX
XX AAG90799;
AC
XX

DT 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 4553.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EP1108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senon A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR
XX N-PSDB; AAH6018.
DR
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
XX Claim 17; SEQ ID NO: 4553; 246bp + Sequence Listing; English.
XX
XX

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX
SQ Sequence 511 AA;

Query Match 5.0%; Score 131.5; DB 22; Length 511;
Best Local Similarity 22.5%; Pred. No. 0.0076;
Matches 104; Conservative 51; Mismatches 199; Indels 109; Gaps 21;

QY 29 KAAQKSSTVDAATAKANAASOEHOGE--LPVIDAIYTHAEVPPVD--RDHPAK 83
DB 11 KAGAVLAATVVAQVIVACSSSDVAGXGGEPTLPI-----PPADLTREGSS 58
QY 84 VVVKMETVEKVRMLADGVEYQWTFGGQVPGQMIRVREGDTIEVQSNHPS----- 135
DB 59 VHFALAEAGQGESQILPDVTTKTWGFNGTHIGPTLVKKKGDVHVIVNNLDEMTVHMHG 118
QY 136 -GMPPNVDP--HAAGPGGGAASFTAPGHTSTFSFKALQPLVYHCAVAPGMHANG 192
DB 119 MRLPAIADGPRSPSPGQWTSPWTVANDAATLTVHPHHTGL-----TGLHAYRG 169
QY 193 MYGLIIVEPEKGLPKVD--KEYYV-----MQGDFYTKGKYGEGLOPFDMKAIRED 243
DB 170 LAGMITIVE-DEATDLDLPREYGVDDIPVLMHDFLEBGSIDDEDLP--DGLGLDPTT 226
QY 244 EYVFNQSVGALTGNNALKAKVGETVRLFVNGGNGNL-----TSSFHVIG-----EIFD 292
DB 227 ANGITNAHFDAATTRVRFRVLNGSNMRFY-----NLAFSDRTFTQVLAISGLDDEPD 280

QY	293	KVHEGGKGENHNHQTLLIPAGGAALT-----EFKVDVPGGVYLVDAIFRAAFKALG	346
Db	281	RTTLAIPGERWEIWLLEP--GEDVTLSVGEEDYGVPPDEFPDPGMSDSFQ----- <td>333</td>	333
QY	347	ILKVEGSENEHIYSHKQTDVAVLP-----EGAFQALDTQEP--KTPAPANLQEQI <td>395</td>	395
Db	334	LLRTTGGSD-----DAAQAPALPGVLKSTEDVIDATERFIMNTESINDLQWDM <td>386</td>	386
QY	396	KAGKATYDSNCACHOPDGKGVNAPFPPLANSDYLNADARAA <td>438</td>	438
Db	385	QRVDVVIDHD-----QPEWIVITN-----DNSDWHNFHVDA <td>417</td>	417
RESULT 36			
ID	AA879105	standard; Protein; 497 AA.	
AC	AA879105;		
DT	30-APR-2001	(first entry)	
XX	Corynebacterium glutamicum HA protein sequence SEQ ID NO:166.		
DE			
XX	Corynebacterium glutamicum; homeostasis; adaptation; HA protein;		
KW	fine chemical production; organic acid; proteinogenic amino acid;		
KW	nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;		
KW	nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;		
KW	carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;		
KW	diagnosis; Corynebacterium diptheriae; genetic engineering;		
KX	Brevibacterium; environmental condition.		
XX	Corynebacterium glutamicum.		
OS			
XX	WO200100842-A2.		
PN			
XX	04-JAN-2001.		
PD			
XX	23-JUN-2000; 2000WO-IB00911.		
PF			
XX	25-JUN-1999;	99US-0141031.	
PR	08-JUL-1999;	99DE-1031636.	
PR	09-JUL-1999;	99DE-1032125.	
PR	09-JUL-1999;	99DE-1032126.	
PR	09-JUL-1999;	99DE-1032127.	
PR	09-JUL-1999;	99DE-1032128.	
PR	09-JUL-1999;	99DE-1032129.	
PR	09-JUL-1999;	99DE-1032226.	
PR	14-JUL-1999;	99DE-1032920.	
PR	14-JUL-1999;	99DE-1032922.	
PR	14-JUL-1999;	99DE-1032924.	
PR	14-JUL-1999;	99DE-1032928.	
PR	14-JUL-1999;	99DE-1032930.	
PR	14-JUL-1999;	99DE-1032933.	
PR	14-JUL-1999;	99DE-1032935.	
PR	14-JUL-1999;	99DE-1032973.	
PR	14-JUL-1999;	99DE-1033002.	
PR	14-JUL-1999;	99DE-1033003.	
PR	14-JUL-1999;	99DE-1033005.	
PR	14-JUL-1999;	99DE-1033006.	
PR	31-AUG-1999;	99DE-1041378.	
PR	31-AUG-1999;	99DE-1041379.	
PR	31-AUG-1999;	99DE-1041390.	
PR	31-AUG-1999;	99DE-1041391.	
PR	03-SEP-1999;	99DE-1042068.	
XX	(BADI) BASF AG.		
XX	Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;		
XX	WPI; 2001-061974/07.		
DR	N-PSDB; AAF71220.		
XX			

[illegible]

XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134421.
PR 14-MAY-1999; 99US-0134570.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

```

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 16-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 5.0%; Score 130; DB 21; length 554;

Best Local Similarity 22.4%; Pred. No. 0.012; Matches 104; Conservative 59; Mismatches 184; Indels 118; Gaps 23;

```

QY 82 AKVVVKME-----TVEKVRMLADGVEYQFTFGQVPGQMIRREGDTIEVOFSN 131
DB 17 ATVVVAEDPYPFHVWNVTVGTSP--GVPOVILINQFPGPNVNSTNNNVIINVEN 74
QY 132 HPD-----SKPHNVDFHAATPGGGAASFTAPGHTSTFSFKL-QPGLVYHCAYA 183
DB 75 NLDEPFLITWNGIOHRKNCMODGTPTMCP--IMPGTNYTHFQPKDDIGSFYYPSTA 131
QY 184 PVGMHIANMGYGLIVERKEGLP---KVDKEYVVMQGFYTKRGKYGEGQLOPF-DMEXA 238
DB 132 ---NHRASGFGGLRVNSRLIIPVYADPEDDTYVILIGWYTK--SHTQKKFLDSGRT 185
QY 239 I-REDAEYVVFNGSVGALTGENA--LKAKVGETVRLFVNGGPNLTSSPHVIGEIFDKVA 295
DB 186 LGRDGS--ILINGSGKGDSDADLPFLKPGKTYRVAICNVGLKTSINLFRIOHKLIVE 243
QY 296 FEGGKGNHNIQTLIPAG--GAIT--EFKVDVPGDYVLDHAI FRAENKALGILK 349
DB 244 MEGSHVLQNDYDSLIVAVGQCQYGTILITANDBAK-----DYVMVASRFLKSVITTTGLR 298
QY 350 VEGEENHEIYSHKQTDVAVLPEG-----AQALDTQDEAPRTAPAN 390
DB 299 YEG-----GKPASSQLPPGPGVGMASLNOFRSFRMLTASAAPNPGSYHYGKIN 350

```

```

QY 391 LOEOLKA-----GKATYDSNCACHOPDKGVNAPFP--PLANSDYLMNDHARAASI 440
DB 351 ITRITKLVTNQGKVDKLRALN-----GVSHTDPEPFLKLAEXF----- 390
QY 441 VANGLSGKITVNGNOYESV---MPAIALSDQIINAVITYTLNSF 481
DB 391 ---GVADKVF---KYDSITDNPTPEQIKSIKIVPNVILITHRTF 428

```

RESULT 38

```

ABB92295
ID ABB92295 standard; Protein; 554 AA.
XX
XX ABB92295;
AC
XX
XX 31-MAY-2002 (first entry)
DT
XX
XX Herbicidally active polypeptide SEQ ID NO 1506.
DE
XX
XX Herbicidal; plant; agriculture; herbicide.
KM
XX
XX Arabidopsis thaliana.
OS
XX
XX WO200210210-A2.
PD
XX
XX 07-FEB-2002.
PF
XX
XX 28-AUG-2001; 2001MO-EP09892.
PR
XX
XX 28-AUG-2001; 2001MO-EP09892.
PA
XX
XX (PAB) BAYER AG.
PI
XX
XX Tietjen K, Weidler M,
DR
XX
XX WPI; 2002-269010/31.
XX
XX
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
XX
XX Claim 5; SEQ ID NO 1506; 261pp + Sequence Listing; English.
PS
XX
XX
XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
XX
XX Sequence 554 AA;
SQ

```

Query Match 5.0%; Score 130; DB 23; length 554;
 Best Local Similarity 22.4%; Pred. No. 0.012; Matches 104; Conservative 59; Mismatches 184; Indels 118; Gaps 23;

```

QY 82 AKVVVKME-----TVEKVRMLADGVEYQFTFGQVPGQMIRREGDTIEVOFSN 131
DB 17 ATVVVAEDPYPFHVWNVTVGTSP--GVPOVILINQFPGPNVNSTNNNVIINVEN 74
QY 132 HPD-----SKPHNVDFHAATPGGGAASFTAPGHTSTFSFKL-QPGLVYHCAYA 183
DB 75 NLDEPFLITWNGIOHRKNCMODGTPTMCP--IMPGTNYTHFQPKDDIGSFYYPSTA 131
QY 184 PVGMHIANMGYGLIVERKEGLP---KVDKEYVVMQGFYTKRGKYGEGQLOPF-DMEXA 238
DB 132 ---NHRASGFGGLRVNSRLIIPVYADPEDDTYVILIGWYTK--SHTQKKFLDSGRT 185
QY 239 I-REDAEYVVFNGSVGALTGENA--LKAKVGETVRLFVNGGPNLTSSPHVIGEIFDKVA 295

```


Db 186 LGPDDG--LLINKSGKSGSDAPLFTLKPKTYRVARICVNGKTSINFRIOHMKLVE 243
Qy 296 FEGKGENNIQTLIPAG---GAAT---EFYVDVPGDVLVDHAIIFRANFGALGILK 349
Db 244 MEGSHVLQNDYSLDVHVGOCYGTIITANEAR-----DYMWASSRFLKSVTTTGLR 298
Qy 350 VEGEENHEIYSHKQDAVYLPEG-----APQAIIDTQEARCTAPAN 390
Db 299 YEG-----GKPPASSQLPPGPGVWAMSLNQFRSFRWMLTASALRPNGSGYHVKIN 350
Qy 391 LGEQITRA-----GKATDSNCAACHQPDGKGVNAPF--PLANSDYINADHARAASI 440
Db 351 IRTTILVNTQGVKDKLRALN-----GVSHDPEPLKAEYF----- 390
Qy 441 VANGLSGKITVNGNOYESV---MPAIALSDQOIANVITVTLNSF 481
Db 391 ---GVADKVF---KIDSTIDNPTPEQISIKIVPNVLTHTRTF 428

RESULT 39

AAB03533
ID AAB03533 standard; protein; 2183 AA.

XX AC AAB03533;

XX DT 09-OCT-2000 (first entry)

XX DE Murine factor V SEQ ID NO:5.

XX KM Murine; factor V; FV; activated protein C; APC; anticoagulant;
XX KM activated protein C resistant factor V; thrombosis; screening;
XX KM thrombophilia.

XX OS Mus sp.

XX PN US6066778-A.

XX PD 23-MAY-2000.

XX PF 06-NOV-1996; 96US-0746111.

XX PR 06-NOV-1996; 96US-0746111.

XX PA (UNMI) UNIV MICHIGAN.

XX PI Ginsburg D, Cui J;

XX DR WPI; 2000-410682/35.

XX DR N-PSDB; AAA60446.

XX PT New transgenic mice expressing activated protein C resistant factor V
XX PT as models for human thrombophilia and as models for testing in utero
XX PT gene therapy protocols -

XX PS Example 1; Fig 2; 76pp; English.

XX CC The present invention describes transgenic mice (I) and (II) containing
XX CC modifications in the factor V gene, where (I) expresses an activated
XX CC protein C (APC) resistant factor V and (II) lacks the ability to express
XX CC wild-type factor V. The transgenic animals (I) and (II) are useful for
XX CC screening compounds with anticoagulant activity. Methods from the present
XX CC invention, and the transgenic animals, are also useful in providing
XX CC models for human thrombophilia. These models are useful in providing
XX CC insight into the basic regulatory mechanisms of blood coagulation and
XX CC pathogenesis of human thrombosis. In addition, factor V null transgenic
XX CC mice, especially pregnant females may be used as a model system to test
XX CC in utero gene replacement therapy protocols. The present sequence
XX CC represents murine factor V, which is used in an example from the present
XX CC invention.

XX SQ Sequence 2183 AA;

Query Match 5.0%; Score 130; DB 21; Length 2183;
Best Local Similarity 20.7%; Pred. No. 0.091;
Matches 70; Conservative 46; Mismatches 120; Indels 102; Gaps 16;

Qy 63 DALVTHAEVPPVDR-----DHPKAVVVKMEVEYVYMLADGEVQFWT 107
Db 357 EYIWNAPVPIPANMDKIVRSOHLDNFSNOIGKKYKVIROYEEFFTRTDPISIK--- 413
Qy 108 FGGQVVGOMIRREGGTVQPSN---HPDSKMPHNVDFHAATGPGGAASFTAPGHNS 164
Db 414 -QSGILGPVIRACVQRTDLTKIVKRNASRPYSIYPHCVTSPYE--DGINSSTGSSHT- 468
Qy 165 TFSFKALPG-LVYV-----HCAVAP-----VGMHIANGMVGLILVEPK 202
Db 469 --TIRPVQGEFFTYKMNILFEDEPTENDACLTIRYSYDVDTVRDIAGLIGLLICKS 526
Qy 203 EGLPK-----VDKEYVMQGF-YTKGKRYGEGLOPF--DNEKAIREDAEVYVFGSVG 253
Db 527 RSLDQGVQVADIEQAVFAVFDEKSWYIEDNINIKFCENPDEVKDDPKF----- 578
Qy 254 ALTGENALAKAGETVRLFVNGGPRULTSSFVIGEIFD-----KVP 296
Db 579 -----YESNINSTINGIV-----PESISTLGCDFDPTVQMHFCSVGTHDDILTIHF 624
Qy 297 EGK---GENNIIQTLIPAGAAITPEFVDPGVLYL 331
Db 625 TGHSFIVGRHEDTJLTFPMRGESVT-VTMADVGTWML 661

RESULT 40

ABP78252
ID ABP78252 standard; protein; 439 AA.

XX AC ABP78252;

XX DT 07-MAR-2003 (first entry)

XX DE N. gonorrhoeae amino acid sequence SEQ ID 3034.

XX KM Antibacterial; infection; vaccine; gene therapy.

XX KM Neisseria gonorrhoeae.

XX OS Neisseria gonorrhoeae.

XX PN MO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-1B02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Piazza M, Masiagnani V, Monaci E;

XX DR WPI; 2003-058415/05.

XX DR N-PSDB; ABZ39222.

XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX PT medicament for treating or preventing N. gonorrhoeae infection -
XX PS Disclosure; Page 410; 815pp; English.

XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX CC antibodies that specifically bind to the proteins. The composition
XX CC comprising the protein, nucleic acid or antibody is useful for the
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
XX CC infection, this may be in the form of a vaccine or gene therapy.
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX CC molecules of the invention.

SQ Sequence 439 AA;
Query Match 4.9%; Score 129.5; DB 24; Length 439;
Best Local Similarity 22.6%; Pred. No. 0.0092;
Matches 79; Conservative 47; Mismatches 131; Indels 93; Gaps 18;
QY 184 PVGNHIANGMVGLIVE-----PKEGLPKVDKEYVMOGD-----FYTGGK----- 224
133 POKQIAQNLFTNYCICQCHGSDAKSGKFPNLTDSDWLWGGDDPKIHETIEKGRVATMPA 192
QY 225 -----YGEQGLQPFMEKAIREDAYV-FNGSVGALTGENALAKV---GETVRLFYGNG 276
193 WGPALGEEG-----VXDVAHYVMSLSKPKQYDEERAARQALFSGSPANCFYCHG 243
QY 277 -----GPNLTSSPFIHGEIFDKVHFEKGKGEN-----HNIQTLLIPAGAAITEFK 322
244 DKGOGTQGLGNLT-----DQVWLMGGTOKSIIETITNGRSSQMPAMGHFLDKX 293
QY 323 VDVPGDYVLVDHAFRAFNK-GALGILKVEGENHEIYSHKQTDVYLLPEGAPQIDTQE 381
294 LHIWTAIV-----WGLSNKDGKAPVKKAEPAPAEAPAPAPAAQAASEAKPAAAEPK 347
QY 382 APKTPAPANLOEQIKA-GKATYDSCAACHQPDGKGVNAPFPPLANDYLNADHAPASI 440
348 AEEKAAPA-----AKADGKQVETVCAACH---GNALPGIPIHVGTKADW--ADRIK----- 393
QY 441 VANGLSGKITVNGNOYE--SVMPAIA-----LSDQIANTVITYTLNDFGNK 484
394 -----KCKDTLHKHAIETGNTMPAKGGRGLSDDEVAADVIMVNGSGK 438
Db

Search completed: August 27, 2003, 18:29:33
Job time : 88 secs

Handwritten: 9/14/1999 Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 5, 2003, 05:17:53 / Search time 407 Seconds
(without alignments)
3329.528 Million cell updates/sec

Title: US-10-088-045-2

Sequence: 1 MSKPTLIKTTTLCALSLML.....NKGQLSADVAKAKTKEN 502

Scoring table: BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 segs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seg length: 0
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DRV=xlh
-Q=/cgr2_1/USFTO.spool/US10088045/runcat.04092003.083141.3908/app_query.fasta_1.647
-DB=N-Geneseq.150un03 -QFMT=fastap -SUFTX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -STRT=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=5 -DOCALLIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORML=0 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10088045 @CCN 1.1.312 @runcat.04092003.083141.3908 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGIOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N-Geneseq.150un03.4*
1: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDS1/gcgdata4/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2626	100.0	1506	22	AAFS5657
2	2626	100.0	1509	22	AAFS5658
3	2626	100.0	66986	22	AAFS28542
4	1394	53.1	349980	21	AAFS21611
5	1394	53.1	837096	21	AAFS1489
6	1388	52.9	1176	25	AB238597
7	1346.5	51.3	7919	21	AAAB81536
8	1034	39.4	691	24	AAI46493
9	349	13.3	37337	23	AAFS5918
10	237	9.0	399	21	AAZ53036
11	237	9.0	552	25	AB239100
12	222	8.5	396	21	AAZ53038
13	222	8.5	402	21	AAZ53037
14	222	8.5	12893	21	AAA81733
15	222	8.5	349980	21	AAFS21608
16	222	8.5	1437668	21	AAAB81490
17	184	7.0	341	21	AAZ53034
18	184	7.0	341	21	AAZ53035
19	175	6.7	4290	23	AB152917
20	172	6.5	342	21	AAZ53033
21	164	6.2	1437	14	AAQ46687
22	163	6.2	1437	14	AAQ48234
23	163	6.2	1437	20	ABA96893
24	163	6.2	1437	21	AAAC83154
25	163	6.2	2064	12	AAQ14178
26	163	6.2	4830	20	ABA96894
27	163	6.2	4830	21	AAAC83156
28	162	6.2	2007	22	AAFS26419
29	158.5	6.0	4403765	22	AAI99683
30	158.5	6.0	4411529	22	AAI99682
31	155	5.9	5187	20	AAZ57912
32	152.5	5.8	1125	25	ABT14589
33	152	5.8	1404	13	AAO20384
34	150	5.7	1404	12	AAQ13581
35	146.5	5.6	1668	12	AAQ12068
36	142	5.4	1533	22	AAH66018
37	142	5.4	1614	22	AAFS71220
38	142	5.4	349980	22	AAH68527
39	141	5.4	1588	18	AAI63317
40	141	5.4	1588	21	AAZ24235
41	141	5.4	1672	16	AAQ86526
42	140.5	5.4	1479	22	AAH68245
43	140.5	5.4	309400	22	AAH68534
44	139	5.3	1722	24	AAI13870
45	136.5	5.2	980	21	AAH68115

ALIGNMENTS

RESULT 1	AAFS5657	AAFS5657 standard; DNA; 1506 BP.
ID	AAFS5657	
XX	AAFS5657	
AC	AAFS5657	
XX		
DT	11-JUN-2001	(first entry) :
XX		
DE	Nucleotide sequence of a BASB109 polypeptide.	
XX		
KM	BASB109; bacterial infection; vaccine; genetic immunisation; ss.	
XX		
OS	Moraxella catarrhalis.	
XX		
Key	Location/Qualifiers	
FT	1..1506	
CDS	/*tag= a	
FT		

```

FT      /product= "BASB109"
FT      /note= "no termination codon given"

```

PN WO200119996-A1.

PD ~~22-MAR-2001~~

PF 14 SEP 2000; 2000WO-EP09035.

PR 14-SEP-1999; 99GB-0021691.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Thomnard J.; PI

DR WPI; 2001-257883/26.

1999

PT Novel BASH09 polypeptides of *Moraxella catarrhalis* is useful for diagnostic, prophylactic and therapeutic purposes against microbial diseases, preferably bacterial infections -

PS Claim 13; Page 65-66; 93pp; English.

The present sequence encodes a BASH109 polypeptide of Moraxella catarrhalis. BASH109 polypeptides and polynucleotides are useful for treating bacterial infections, and as research reagents and materials for the treatment of and diagnosis of diseases, particularly human diseases. They are useful for inducing an immune response in an individual, and to assess the binding of small molecule substrates and ligands in, for e.g., cells, cell-free preparations, chemical libraries, and natural product mixtures. BASH109 polynucleotides are useful for therapeutic or prophylactic purposes, in particular genetic immunisation and in diagnosis of the stage and type of infection.

Sequence 1506 BP; 431 A; 356 C; 363 G; 356 T; 0 other;

Alignment Scores:

Pred. No.:	2,52e-227	Length:	1566
Score:	2626.00	Matches:	502
Percent Similarity:	100.00%	Conservative:	0
Beet Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-10-088-045-2 (1-502) X AAF55657 (1-1506)

1 MetSerLysProThrLeuIleLysThrTh

Db 1 ATGCTAGCCTACTTTGATAAAACAAC

21 SerGlyCySerAsnGlnAlaAspLysAla

Db 61 AGTGGTTGTAGCAATCAAGCGGACAAAGCC

US-10-088-045-2 (1-502) X AAF55657 (1-1506)

Cy 501 Proben 502
|||||
Db 1501 CCNAC 1506

RESULT 2
ID AAF55658 standard; DNA; 1509 BP.
AAFS5658
AC AAF55658;
AT 11-JUN-2001 (first entry)
DT Nucleotide sequence of a BASB109 polypeptide.
DE
XX BASB109; bacterial infection; vaccine; genetic immunisation; ss.
XX Moraxella catarrhalis.
OS
XX Key Location/Qualifiers
PH 1..1509
FT CDS /*tag= a
FT /product= "BASB109"
XX MOZ0UT119996-1.
XX PD 22-MAR-2001
XX PF 14-SEP-2000; 2000MO-EP09035.
XX PR 14-SEP-1999; 99GB-0021691.
XX PA (SMIK) SWITKLINE BECHAM BIOLOGICALS.
XX PI Thomard J;
DR WPI; 2001-257883/26.
DR P-PSDB; AAB67668.

Novel BASB109 polypeptides of Moraxella catarrhalis useful for diagnostic, prophylactic and therapeutic purposes against microbial diseases, preferably bacterial infections -
Claim 13; Page 65; 93pp; English.

The present sequence encodes a BASB109 polypeptide of Moraxella catarrhalis. BASB109 polypeptides and polynucleotides are useful for treating bacterial infections, and as research reagents and materials for the treatment of and diagnosis of diseases, particularly human diseases. They are useful for inducing an immune response in an individual, and to assess the binding of small molecule substrates and ligands in, for e.g. cells, cell-free preparations, chemical libraries, and natural product mixtures. BASB109 polynucleotides are useful for therapeutic or prophylactic purposes, in particular genetic immunisation and in diagnosis of the stage and type of infection.

SQ Sequence 1509 BP; 432 A; 356 C; 364 G; 357 T; 0 other;

Alignment Scores:
Pred. No.: 2.53e-227 Length: 1509
Score: 2626.00 Matches: 502
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-088-045-2 (1-502) x AAF55658 (1-1509)

Cy 1 MetSerLysProThrLeuIleValThrThrLeuIleCysAlaLeuSerAlaLeuMetLeu 20
Db 1 ATGCTCAAGCTACTTTGGATTAAACAACCTTAATTGTGCCCTTAAAGTGATGATGCCTC 60
21 SerArgCysSerAsnGlnAlaAspPheAlaAlaGlnProLysSerSerThrValAspAla 400

Db	61	AGTGGTTTACCAATCAACGCCGACAAACCCGCCACCAAAAAGACACGGTAGACGCT	120
Oy	41	AlaAlaIysThrAlaAsnAlaAspAsnAlaAlaSerGlnIuHISGlnIyGluLeuPro	60
Db	121	GCCGCCAAGACAGCAAAATGACAGATATATCTGCTCACAAGAACATCAAGGCGAGCTGCT	180
Oy	61	VallIleAspAlaIleValThrHisAlaIleProGluValIleProProValAspArgAspHis	80
Db	181	GTCAATGATGCGCATTTGTCATGCAATGACCAAGAGTTCCACCACTGTTGACCGTGACCAAC	240
Oy	81	ProAlaIysValIleValIleMetGluThrValGluIysValIleArgLeuAlaAspGly	100
Db	241	CCCCGCCAAGTGGTGTAAAAATGGAACCGTTGAAAAAGTCATGCGTCTGGCAGATGGC	300
Oy	101	ValGluIyGlnIlePheIyPhePheGlyGlyGlnValIleProGlyGlnMetIleArgValArg	120
Db	301	GTGGAAATATCAGTTTGTGGACATTTGGCGCGGTCAAGTTCCAGGGCAGATGATGTTGTTGGCGCT	360
Oy	121	GluGluIysPheThrIleGluValGlnPheSerAsnHisProAspSerIleMetProHisAsn	140
Db	361	GAAGCGCAGACCATCGAAGTGCAGTCTCCAAACCCACAGATTCAAAATGGCCCATATAT	420
Oy	141	ValAspPheHisAlaIleThrGlyIleProGlyGlyValIleGluAlaSerPheThrAlaPro	160
Db	421	GTGACCTTTCACGCTGCCACAGGGCTTGGCGGGGAGAGAGCGTCAATTTACCGCACCG	480
Oy	161	GlyHisThrSerThrPheSerPheIysAlaLeuGlnProGlyLeuIyValIyHisCys	180
Db	481	GGTCATACATCAACCTTTAGTTTAAAGCCTTACAGCGCTGGTTGTATGTCATCATCTGCT	540
Oy	181	AlaValAlaIleProValGlyMetHisIleAlaAsnGlyMetIyGlyLeuIleLeuValGlu	200
Db	541	GCGGTGGCCCTGGTGGCATGCACATTCCTATGCAATGCAATGATGGTTGATTTGGTTGAA	600
Oy	201	ProIyGlnGluIyLeuProIyValIleAspIyGluIyIyValIleMetGlnIyAspPheIyArg	220
Db	601	CCAAAAGGGGCTTGGCCAAAAGTAAAGATACATATGTCATGCCAAGGCCACCTTTAT	660
Oy	221	ThrIyGlnGlyIyIyGlnIyGlnIyLeuGlnProPheAspMetGluIyValAlaIleArg	240
Db	661	ACCAAGGCCAATATGGCGAACAAGTCTACAGCCCTTTGATATGGAAAAACCAATTCGA	720
Oy	241	GluAspAlaGluIyIyValIlePheAsnGlySerValIyAlaIleuThrGlyGluAsnAla	260
Db	721	GAAGATGCTGAATATGTTGCTTAAATGTTGCTGGTGGGGGCGCTTCACTGGTAAAAATGCT	780
Oy	261	LeuIyValIyIysValIyGluIyThrValArgIleuPheValIyIyAsnGlyGlyProAsnIleu	280
Db	781	CTAAAGGCCAAGTTGGCGAAACGTGGCTTATTTGGGTGAACGGCGGCCGGAATTTG	840
Oy	281	ThrSerSerPheHisValIleGlyGluIlePheAspIysValHisPheGluIyGlyIys	300
Db	841	ACATCATCATTCCTCATGTCATTTGGTGAAGATTTTGTATTAAGTTCACCTTTGAGGGTGTAAAG	900
Oy	301	GlyGluIleAsnHisAsnIleGlnThrThrLeuIleProAlaGlyIyAlaAlaIleThrGlu	320
Db	901	GGTGAAAACCAATATATCCAAACACCGCTATCCACAGCAGGTGGCGCTGCCATCTGAA	960
Oy	321	PheIyIysValIleAspValProGlyIyAspIyIyValIleuValAspHisAlaIlePheArgAlaPhe	340
Db	961	TTTAAGGTGATGTCGGCGGTGATTAATGTCCTGTTGTCACATGTCATCTTCGTCGCTTT	1020
Oy	341	AsnIyGlnAlaIleuGlyIleLeuIyValIyGluGlyGluIyAsnHisGluIleIyIySer	360
Db	1021	AACAAAGGGCATTTGGGCATCTTAAGGTGAAAGTGAAGAAATCAAGAGATTAATTCGA	1080
Oy	361	HisIyGlnIyThrAspAlaIleValIyIleuProGluGlyAlaProGlnAlaIleAspThrGln	380
Db	1081	CACAAACCAACAGACGCTGTATCTGCGCAAGAGGGTGGCCCAACAGCAATTAATGCCCA	1140
Oy	381	GluAlaIleProIyIyThrProAlaIleAsnIleuGlnIyGlnIleIyValIleGlyIyValAla	400

Dd	4642	GGTAAAAACCATATTCCTCAACACCGTAACTCCAGAGGTGGCGCTGCATCACTGAA	4701
Oy	321	PhelysValaAapValProGlyAspTyrValLeuValaAspHisSalalIlePheAryAlaPhe	340
Dd	4702	TTTAAAGGTGATGTGCGCGGTGATATATGCTTTGGTTGACCAATGCCATCTTCGAGCCTTT	4761
Oy	341	AsnlySGlyAlaLeuGlyIleLeuLysValIuglyIuglyIuamHisGluIleTyrSer	360
Dd	4762	AACAAAGGGGATTTGGCGCATCTTAAGGTGAAAGGTGAAGAAATTCATGACATTTATTTCA	4821
Oy	361	HislySGlnThrAspAlaValTyrIleuProGlyIuglyAlaProGlnAlaIleAspThrGln	380
Dd	4882	CACAAACAAACAGACGCTGTCTATCTGCCAGGGGTGCCCAACAGCAATGATGCCAA	4881
Oy	381	GluAlaProlysthrProAlaProAlaAsnLeuGlnGlnIleLysSalaglyLysAla	400
Dd	4882	GAAGCACCACCAAAACCTCGCACCTGCCCACTTACAGAGCAGATTAAAGCAGTAAAGCA	4941
Oy	401	ThrTyrAspSerAsnCySalAlaAcYHisGlnProAspGlyLysGlyValProAsnAla	420
Dd	4942	ACCATGACTCTTAACCTGTGCTGTTCACCAACTGATGGTAAAGGCGGTGCCAAACGCT	5001
Oy	421	PheProPoleuAlaAsnSerAspTyrIleuAsnAlaAspHisSalAryAlaIleSerIle	440
Dd	5002	TTCCACCGCTTGCCCACTCTGACTCATCTGAACGCCGACCAAGCTCGCGCCAGCAATC	5061
Oy	441	ValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlySerValMet	460
Dd	5062	GTGGCAATGATGATTGTCTGTGAAGTTACCGTCATGGCAACCAATGAAAGGTCATG	5121
Oy	461	ProAlaIleAlaLeuSerAspGlnGlnIleAlaIleAsnValIleThrTyrThrLeuAsnSer	480
Dd	5122	CCTCGATGCTCTGACGCGACCAACAGATTCCTCAATGCATCACCGCTTAACAGC	5181
Oy	481	PheGlyAsnlySGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLysIleThrLys	500
Dd	5182	TTTGTGAACAAAGGCGGTCAACCTGACAGACGATGTGGCAAAAGCCAAAAAAACCAAG	5241
Oy	501	ProAsn 502	
Dd	5242	CCAAAC 5247	
RESULT 4			
AAAF21611			
ID	AAAF21611 standard; DNA; 349980 BP.		
XX	AAF21611;		
XX	AC		
XX	DT		
XX	13-MAR-2001	(first entry)	
DE	Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;		
KW	diagnosis; antigen; detection; infection; gene therapy; antibacterial;		
ds.			
XX	Neisseria meningitidis.		
XX	OS		
XX	WO2000066791-A1.		
PN	09-NOV-2000.		
XX	09-NOV-2000.		
XX	08-MAR-2000; 2000WO-US05928.		
PF			
PR	30-APR-1999; 99US-0132068.		
PR	08-OCT-1999; 99WO-US23573.		
PR	28-FEB-2000; 2000GB-0004695.		
XX			
XX	(CHIR) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
PI	Pizza M, Hickey E, Peterson J, Tettelein H, Venter JC, Masignani V;		

PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R,
PI Frazer CM, Grandi G,
DR WPI, 2000-647603/62.
XX
XX
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PI frames are used to detect, treat and prevent Neisserial infections -
XX
XX Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of
CC *Neisseria meningitidis* B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC *Neisseria* proteins given in AAB58350 to AAB58853, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to *Neisseria* bacteria or as a diagnostic reagent for detecting the
CC presence of *Neisseria* bacteria or of antibodies raised to *Neisseria*
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.

SQ Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 other;

Alignment Scores:	
Pred. No.:	5 66e-113
Score:	134.00
Percent Similarity:	79.74%
Best Local Similarity:	67.95%
Query Match:	53.08%
DB:	21
	Gaps: 2
	Indels: 4
	Mismatches: 75
	Matches: 25
	Length: 349980

US-10-088-045-2 (1-502) X AAF21611 (1-349980)

Qy	1	MetSerIyPProThrIleuIleIySthrIreIleIcQyAlaIleuSerAlaIleuMetIeu	20
Db	187900	TTACAAAAGGAAAAATTATGTAAGACCAAGCCATTAGCTGCAATGATCTTCCTTATTC	1879559
Qy	21	SerGIyCysSerAsnGlnAlaAspLysAlaIaGlnProIySerSerThrValaAspAla	40
Db	187960	GCATTAGCCGCTGCGCGCGGCAACCTGCGCGCAAGCCCTTCGCAAACTCCTGCCCT	1880139
Qy	41	AlaAlaIySthrAla--AsnAlaAspAsnAlaAlaSerGlnIleuIaGlnIyIleu	59
Db	188020	GCCGCGCAAGCCGCAAGCTCGCGCGCAAAACCGCGCGCAACACCGTCGGCGAACTG	1880797
Qy	60	ProValIleAspAlaIleValThrIaIaProGluValProProValaAspAlaAsp	79
Db	188080	CCCGTTATCGATGCGGTATCCACACCCAGCTCCCGAAGTCCCTCTGCAATCGACCCGAG	1881399
Qy	80	HisProAlaIyValaIValIyIleMetGluThrValaGluIyValaIleMetArgIleuAlaAsp	99
Db	188140	TACCCCGCCCAAGTCGCGCTTAATAATGGAACCTCGAATAAACATGACCATGGAAGAC	1881999
Qy	100	GlyValaGluIyGlnPheIyIyPheIyGlnIyValaProGlyIleMetIleArgVal	119
Db	188200	GGTGGGAATACCGCTACTGACATTGACGGCGAGCGTTCGGGCGCGTATGATCCGCGTA	1882599
Qy	120	ArgGlnIyAspThrIleGluValaGlnPheSerAsnHisProAspSerIyMetProHis	139
Db	188260	CGCAAGCGATACCGTGAAGTGAATTTCCAACTCTTCTTCACTGCTTCGCGAC	1883199

OY		140	AanValAlAspPheHisIalAaTRGlyProGlyVGIyGlYAlaGluLysAspPheThrAla	159
Dd		188320	AAcGTCACTTCCACCGGGCTACCAGGCAGGGCGGGCGGGCCGCCAACCTTTACCGCT	188379
OY		160	ProGlyHISerThSeriPheSerPheLysAlaLeuGlnProGlyLeuTyValTYrHis	179
Dd		188380	CCGGGGCGATCTTCACCATTCACACTTCAAAGCCTTGCAACCGGGGTGTGTAATCTAC	188439
OY		180	CysAlaValAlAProValGlyMetHisIIeAlaAsnGlyMeTyrgLyLeuIIeLeuVal	199
Dd		188440	TCCGCGCGTGCACCCGGTCCGTATGCACATCGCCAAACGGTATGTACCGTCTGATTGGTTC	188499
OY		200	GluProLysGluGlyLeuProLysValAspLysGluTyTrTYrValMetGlnIyAspPhe	219
Dd		188500	GAGCTTAAGAAGGCTTCGCCGAAGTGATAAAGATTCTACATCGTCCAAGCGACTTC	188559
OY		220	TYrThrLYSGlyLYserTYrglyGluGlnGlyLeuGlnProPheAspMetGluLYsAlaIle	239
Dd		188560	TACACCAAAAGGCAAAAAGGCGCGCAAGGTCTGCAACCGTTGATATGACAAGCCGTT	188619
OY		240	ArgGluAspAlaGluTyValValIPheAsnGlySerValGlyAlaLeuthrGlyGluAsn	259
Dd		188620	GCCGAAACACCTCGAATACTCGATTCAACCGGTCAAGTAGTGTCTATCCGCCGCGATTAAC	188679
OY		260	AlaLeuValAlalyValGlyGluThrValArgLeuPheValGlyIysnGlyProAsn	279
Dd		188660	GCGCTGAAGGCCAAGACAGCGCAACCTGATGATAGCTTGTGTAACGGCGGTCCGAAC	188739
OY		280	LeuthrSerSerPheHisValIIeglyGluIlePheAspLysValHisPheGluGly	299
Dd		188740	TTGGTATCTTCTCTCCACGTCACTCGCGCAAACTTCGACAAAGTTATGTGAAGCGGC	188799
OY		300	LysGlyGluAsnHisAsnIIegInthrThrLeuIleProAlaGlyAlaIalethr	319
Dd		188800	AAACTGATTAACGAAGAAACCTACAAACACCATGCTCTGCGCGGCTCTGCCATCGTC	188859
OY		320	GluPheLysValAspValProGlyAspTYrValLeuValAspHisAlaIlePheArgAla	339
Dd		188860	GAATTCAAAGTGACATCCCGGGCACCTACCTTGTTGACACCTATCTTCCCGCA	188919
OY		340	PheAsnLYSGlyAlaLeuGlyIleLeuLYsValGluGlyGluGluAsnHisGluIleTYr	359
Dd		188920	TTCAACAAAGCGCGCACTGGGTCAATTGAAGTGAAGGTGCAGAAAACTCTGAATCATG	188979
OY		360	SerHisLYSGlInthrAspAlaValTYrLeuProGluGlyAlaProGlnAlaIleAspThr	379
Dd		188980	ACTCAAAATTAAGTGATACCGCTTACCGCGGTACCGGTGACACTCTCTGCT-----	189030
OY		380	GInGUlaIAProLYSThrProAlaProAla	389
Dd		189031	GCTTCGCTCCCGCAGACTTCTGCCCCGCA	189060
RESULT 5				
AAAB1489				
ID.	AAAB1489	standard; DNA;	837096 BP.	
XX	AAAB1489;			
AC				
DT	04-DEC-2000	(first entry)		
XX				
DE	N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.			
XX				
KW	Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;			
KX	antigen; vaccine; diagnosis; infection; antibacterial; identification;			
XX	Meningococcus B; MenB; db.			
OS	Neisseria meningitidis.			
XX				
PN	WO200022430-A2.			
DD	20-APR-2000.			
XX				
PF	08-OCT-1999;	99WO-US23573.		

```

XX 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
PA
PI Fraser CM, Hickey E., Peterson J, Tetelin H, Venter JC, Maignani V, Galeotti C, Mora M, Ratti G, Scarelli M, Scarlato V,
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarelli M, Scarlato V,
XX Repulli R, Pizza M,
XX WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 629-865; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AA81453 to AA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AA81260 to AA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
CC AA81452 represent Neisseria meningitidis MemB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;
XX
Alignment Scores:
Pred. No.: 1,82e-112 Length: 837096
Score: 1394.00 Matches: 285
Percent Similarity: 79.74% Conservative: 46
Best Local Similarity: 67.95% Mismatches: 75
Query Match: 53.08% Indels: 4
DB: 21 Gaps: 2
XX
US-10-088-045-2 (1-502) x AA81489 (1-837096)
QY 1 MetSerIySProThrLeuIlleYerThrThrLeuIlleCYAlaLeuSezAlaLeuNecIeu 20
Dj 252671 TTACAAAGGAAAATATTATTAAGAACCGCAAGCCTTAGCGCATATTCCTTCTTATTC 252730
QY 21 SerGIyCySSerAenGlnAlaSprLySaAlaGInProLYSerSerThrValAspaJa 40
Db 252731 GCATTAGCGCGGTGGCGGCGGAACCTCGCGCAAGCCCCGCCGAACCCCTCGCGCT 252790
QY 41 AlAlaIalysTrHrAla--ASmAlaAspaSmaAlaAlaSerGInGluIHISelIndLyGluLeu 59
Db 252791 GCCCGCGAAGCGGCAGACTCGCGCGCACAAACC GCCCGCGGAACAACCTCGCGCGAACGTG 252850
QY 60 ProValIIIEApAlaIIIEAlIeThrHisAlaPProGluValIProProValAspaPaSp 79
Db 252851 CCCGTTATCGATGCGGTTCACACCCAGCCTCCCGAAGTGTCCTCTCGCATTCACCGCGAC 252910
QY 80 HisProAlaIalysValIalValIalysMetCIuThrValIGluIyValIeetArgIeulaeap 99

```

```

DB      252911  TACCCCGCAAGTCCGCTAAATAAGAAACCGTGAAGAAACCACTGACCAATGAAAGAC 252970
OY      100  GYValAGluTyrGlnPheThrPheGlyGlyValProGlyGlnIleLeuVal 119
DB      252971  GGTGTGAATACCGCTACCTGACATTGACGGGACGCTCCGGCCGTTATGATCCGCTA 253030
OY      120  ArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139
DB      253031  CCGAAGCGCATACCGTTGAAGTGAATTTTCCAAACAATCTTCTTAACCGTCCGAC 253090
OY      140  AsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluAspPheThrAla 159
DB      253091  AACGTGACTTCACACCGCTCCACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 253150
OY      160  ProGlyHisThrSerThrPheSerPheLysAlaLeuGluProGlyLeuValTyrHis 179
DB      253151  CCGGCGGCTACCTCCATTCACATTCCTCAAGCCCTCAACCGGCTCTGATCACTACAC 253210
OY      180  CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuVal 199
DB      253211  TCGCGCGTGGCACCGCTCGGTATGACATCGCAACGGTATGACGCTGATTTTGGTC 253270
OY      200  GluProLysGluGlyLeuProLysValAspLysGluTyrTyrValMetGlnGlyAspPhe 219
DB      253271  GAGCTTAAAGAGGCTCCGCAAGGTGAAGTAAAGATTCTACATCGTCAAGGCGACTTC 253330
OY      220  TyrThrLysGlyLysTyrGlyGlnGlyLeuGlnProPheAspMetGlnLysAlaIle 239
DB      253331  TACACCAAAAGGCAAAAGGCGGCGGCAAGGTCTGCAACCGTCCATATGGAACAAAGCGGT 253390
OY      240  ArgGluAspAlaGluTyrValValPheAsnGlySerValGlyValLeuThrGlyGluAsn 259
DB      253391  GCGGAACACCTGAATACCTCGATTCACACGTCACGTAGTGTCTATCCCGCGCATAC 253450
OY      260  AlaLeuLysValLysValAlaGlyLeuThrValArgLeuPheValGlyAsnGlyLeuProAsn 279
DB      253451  GCGCTGAAGAGCAAGGCGGCAAGCTGATGATGATGATGATGATGATGATGATGATGATG 253510
OY      280  LeuThrSerSerPheHisValIleGlyGlyIlePheAspLysValHisPheGluGlyGly 299
DB      253511  TTGGTATCTCTCTCCACGTCATCGGCAAAATCTTGACAAAGTTATGTTGAAGCGGCG 253570
OY      300  LysGlyLysAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValAlaIleThr 319
DB      253571  AATCGATTACGAATAACGTACAAACACCATGCTTCCTCGCGGCTCTGCGCATCTC 253630
OY      320  GluPheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAla 339
DB      253631  GAATTCAAAGTGCACATCCCGGCGACCTACATTTGTTGACACATCTATCTTCGCGCA 253690
OY      340  PheAsnLysGlyValAlaLeuGlyIleLeuLysValGluGlyGluAsnHisGluIleTyr 359
DB      253691  TTCACCAAGGCGCACTGGGTCAATTGAAGTAGAAGGTCAGAAAAACCTGAAACATG 253750
OY      360  SerHisLysGlnThrAspAlaValTyrLeuProGlyGlyValAspGlnAlaIleAspThr 379
DB      253751  ACTCAAAATTTAGATACATCCGCTTACCGCGGTAACGGTGCAGCTCTCTGCT----- 253801
OY      380  GlnGluAlaProLysThrProAlaProAla 389
DB      253802  GCTTCGCTCCGCGACGCTTCTGCGCGGCA 253831

```

RESULT 6
AB238597
ID AB238597 standard; DNA; 1176 BP.

XX AB238597;

XX 07-MAR-2003 (first entry)

XX N. gonorrhoeae nucleotide sequence SEQ ID 1783.
XX

```

KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
XX Neisseria gonorrhoeae.
OS
XX WO200279243-A2.
XX
XX 10-OCT-2002.
XX
XX 12-FEB-2002; 2002WO-1B02069.
XX
XX 12-FEB-2001; 2001GB-0003424.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Fontana MR, Pizza M, Malignani V, Monaci E;
PI
XX MPI: 2003-058415/05.
XX
XX P-PSDB; ABP7627.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection -
XX
XX Disclosure; Page 316; 815pp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records AB237706-AB242016 represent nucleic acid
XX molecules of the invention.
XX
XX Sequence 1176 BP; 285 A; 371 C; 299 G; 221 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 9,47e-116 Length: 1176
XX Score: 1388.00 Matches: 265
XX Percent Similarity: 81.05% Conservative: 43
XX Best Local Similarity: 69.74% Mismatches: 66
XX Query Match: 52.86% Indels: 6
XX DB: 25 Gaps: 3
XX
XX US-10-088-045-2 (1-502) x AB238597 (1-1176)
XX
OY 11 LeuIleCysAlaLeuSerAlaLeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAla 30
DB 25 ARGATTGCTTCTTATTCGCAATGGCCGCTGCGGC-----GGCGAACCCTGCC 72
OY 31 AlAGlnProLysSerSerThrValAspAlaAlaIleThrAla--AsnAlaAspAsn 49
DB 73 GCGCAAGCCCTCGCGAAACCCCTCCGCTCCGCAAGGCGGCAAGGCGGCAAGTCCGCCACAA 132
OY 50 AlaAlaSerGlnGlnIleGlnGlyLeuProValIleAspAlaIleValThrHisAla 69
DB 123 GCTACCGCGCAAGGCGTGAAGGCGGCAAGTCCGCTGATGATGATGATGATGATGATGATG 192
OY 70 ProGluValProProProValAspArgAspHisPProAlaLysValValValLysMetGlu 89
DB 193 CCGGAAGTACCTCCCGCAATCGACGCGCATATCTCTGCCAAAGTACGGGTAAATGGA 252
OY 90 ThrValGluLysValMetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGly 109
DB 253 ACCGTGAATAACCATGAATAAGTGAACAGGAGGTGAATACCGCTACGACATTTGAC 312
OY 110 GlnGluValProGlyGlnMetIleArgValAlaArgGluGlyAspThrIleGluValGlnPhe 129
DB 313 GCGGACGTTCCGGCGGCGTATGATCCGCGTACCGGAAGGAGTACGATTGAAGTCAATTC 372
OY 130 SerAsnHisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGlyPro 149
DB 373 TCCAAACAATCTCTTCTTACGCTTCCGCAACAGTGCAGCTTCCAGCGGCAACCGGTGAG 432

```

QY		150	GlyGIgIyAlaagluAaserPhePrnAlaProglYHtShSerThrPheSerPhelys	169
Db		433	GGGGGCGGTGCACCCGGCACTTATCCGCCCCCCGGAGCATTCCAACTTCGAATTAAGTTCAA	492
QY		170	AlaLeuGlnProGlYleuTYValTYHtVcYAlaValAlaProValGlYmethisile	189
Db		493	GGCTTCAACCGGGCTGTACTCATCTCACTCGGCGCGTGCAGTATGCACATC	552
QY		190	AlaAsnGlmetyrcdyleuileauValGluProlyGluGlyleuProlysValaAp	209
Db		553	GCCACGGATGTACGGTCTGATTTGGTCGAGCTTAAAGAAGGCTTGCGGAAGTGAT	612
QY		210	LysGluTYrTYValMetGlnLYaerPheTYrThLysGlyLYeSTyrGlyGluGlnGly	229
Db		613	AAAGAGTTCTACATCTGCCAAGCGCACTTCTACACCMAAGCAAAAAGCGCCGCAAGCG	672
QY		230	LeuGlnProPheaSerMetGluLyalaallearGluasparaGluTYrValYPheaen	249
Db		673	CTGCACAACCGTTGGATGTGACAAAGCCGTTCGCCAACACCTCAATACGTCGATTCAAC	732
QY		250	GlySerValGlyAlaleuthrgGlyuaAsnalaleuLYeValalyValGlyluthrVal	269
Db		733	GGCCACAGTAGGGGCTATCCCGCGGCATTAACGCTTGAAAGCCAAAGCGGAACCGTG	792
QY		270	ArgLeuPheValGlyAsnGlyGlyProaenLeuthrSerSerPheHisValIlleglylu	289
Db		793	CCTAATGATAGTGGGTAAACGGCGCCGAACCTGGATGTTCTTCCACAGTCATCGGGCA	852
QY		290	IlePheArlyValHtshPheGluGlylyuLyGlyGluasnHtasnllegluthrThr	309
Db		853	ATCTTCGACAAAGTTATTTGTAAGCGCGCAAACTGATTTAAAGAAAGCTTAAGACAC	912
QY		310	LeuileProAlaGlyGlyAlaAlaIethrgluPheLYaAspValProGlyAspTYr	329
Db		913	ATCGTGGCTCGCGGGCGTCTCGCATCGCGAATTCAAGTGCAGATCCCGGCAACTAC	972
QY		330	ValLeuValaerHtAlaIlePheArglaPheasnLYeGlyAlaLeuGlylleyeulsy	349
Db		973	ACTTGTGTCGACCACTCCATCTTCGCGCATTCMAAAGGCGCTTGCGCAATTGAAA	1032
QY		350	ValGluGlyGluGluasnHtsgluiletyrSerHtsLYeGlnThrAspAlaValtyrleu	369
Db		1033	GTAAGAGGTGGCGGAACCTGAATCATGACTCAAAAATTGATGATACCGCTTACGCC	1092
QY		370	ProGluGlyAlaProGlnAlaIleaPthrGlnGluAlaProLYsthrProAlaProAla	389
Db		1093	GGCAGCGCGCGGCTTTCGCC---CGTCTGCTTCGCACCGGCTGTCTTGCCCCGCGCA	1149
RESULT 7				
ID	AAA81536/C			
XX	AAA81536	standard; DNA; 7919 BP.		
AC	AAA81536;			
DT	04-DEC-2000	(first entry)		
DE	N. meningitidis partial DNA sequence gnm_83 SEQ ID NO:83.			
XX				
KW	Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;			
KW	antigen; vaccine; diagnosis; infection; antibacterial; identification;			
XX	Meningococcus B; MenB; ds.			
OS	Neisseria meningitidis.			
PN	WO200022430-A2.			
PD	20-APR-2000.			
PF	08-OCT-1999;	99MO-US23573.		
PR	09-OCT-1998;	98US-0103794.		
XX	30-APR-1999;	99US-0132068.		
XX				

PA (CHIR) CHIRON CORP.
XX
XX Frazier CM, Hickey E, Peterson J, Terteljin H, Venter JC,
PI Mesinger V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
PI Rappuoli R, Pizza M,
XX
XX WPI, 2000-318079/27.

Claim 7; Page 1476-1478; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent *Neisseria meningitidis* *MemB* polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

Sequence 7919 BP; 2282 A; 1729 C; 1972 G; 1936 T; 0 other;

Alignment Scores:	
Pred. No.:	6,8e-111
Score:	1346.50
Percent Similarity:	84.94%
Best Local Similarity:	75.30%
Query Match:	51.28%
DB:	21
Length:	7919
Matches:	250
Conservative:	32
Mismatch:	47
Indels:	3
Gaps:	1

US-10-088-045-2 (1-502) x AAA81536 (1-7919)

Oy	58	GlueuProValIleAspAlaIleValThrIslaPProGluValProProValAsp	77
Db	7917	GAAGTCGCCGTTATCGATGCGGTTACACCAACCGTCCCGAAGTCCCTCGCAATCGAC	7855
Oy	78	ArgAspHisProAlaIleValValIleValMetGluThrValGluIleValMetArgLeu	97
Db	7857	CGGCACACTACCCCGCCAAAGTCCGGCTAAATATGAAACCGTCGAAAAAACCATGCATG	7798
Oy	98	AlaAspGluValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetCile	117
Db	7797	GAAGACGGTGTGGAATACCGCTACTGACATTGAACGGACACTTCCGGCCGCTATGATC	7733
Oy	118	ArgValArgGluGluAspThrIleGluValGlnPheSerAsnHisProAspSerIleMet	137
Db	7737	CGCGTACGCCAAGGCGATTACGGTGTGAAGTGGAATTTCCAACAATCTCTTCTTACCGT	7678
Oy	138	ProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyGlyValAlaGluAlaSerPhe	157
Db	7677	CCGCACAAACGTGCACCTTCCACGCGGCTAACCGGACGGGCGGCGCGGCGCAACTTT	7618
Oy	158	ThrAlaProGlyHisThrSerThrPheSerPheIleGlnValLeuGlnProGlyIleuTyrVal	177

```

Db      7617 ACCGCTCCGGGCGCTACTTCCACATTGAGCTTCAAAACCCCTGCAACCGGGTCTGTATCATC 7558
Qy      178  T Y H I S C Y S A I A V A I A P R O V A I G I M E T H I S I L E A S N G I M E T Y R G I Y L E U I L E 197
Db      7557 T A C C A C T G C C G C C G C A C C G G T G T A T G C A C T C C A A C G G T A T G T A C G G T C T A T T 7498
Qy      198  L e u V a I g I u P r o I y S g I u I y L e u P r o I y V a I A s p I y S g I u T y r V a I M e t G I n G I 217
Db      7497 T T G G T C A G C C T A A A G A G C C T G C C A A A G T G A T A A G A G T T C A T C A T G C T C A A G G C 7438
Qy      218  A s p P h e t y r T h r I y S g I y L y r G I y G I n G I n G I n P r o P h e a p M e t G I u L y S 237
Db      7437 G A C T T C T A C C A A A G G C A A A A A G G G C C A A G T G C A A C C G T T G C A T G A T G A C A A A 7378
Qy      238  A l a I l e a r g I u S p a I a g I u T y r V a I V a I P h e a n g I y S e r V a I g I a I a l e u T h r G I 257
Db      7377 G C C G T T C C C A A C A G C C T G A A T A C G T G A T T T C A C G G T C A C G T A G G T G C T A T G C C G G C 7318
Qy      258  G I u A s n A I a l e u I y S a I a y S a I g I y L u T h r V a I A r g u e P h e V a I g I A s n G I y G I 277
Db      7317 G A T A A C G C G C T G A A A G C A A G C A G G C A A A C T G A T G A T G A C G T T G A A C G G C G T 7258
Qy      278  P r o A s n l e u T h r S e r S e r P h e I S V a I I l e G I y G I u I e P h e a S p I y S V a I H I s p h e G I u 297
Db      7257 C C G A A C T T G T A T C T C T C C A C G T A T C G G C A A A T C T T C G A C A A A G T T A T G T T A G A 7198
Qy      298  G I y G I y L y S g I y L u S a n I l e G I n T h r I l e u I l e P r o I a g I y G I A l a I A 317
Db      7197 G G G G C A A A C T G A T T A A C G A A A A C G T A C A A G C A C C A T C C T G C C G G C G G T G C C 7138
Qy      318  I l e T h r G I u P h e I y S a I a P r o I y A s p I y r V a I l e u V a I A s p I a I a I l e P h e 337
Db      7137 A T C G T G A A T T C A A A G C A C A T C C G G C A G C A C T T G G T G A C C A C T C A T C T T C 7078
Qy      338  A r g A I a P h e a n I y S g I y A l e u G I y I l e u I y S a I g I u G I u A s n I S g I u 357
Db      7077 C G G C A T T C A A A A G C A C G C A C T G G T C A A T T G A A G A A G T G C A G A A A C C C T G A A 7018
Qy      358  I l e T y S e r I a S y S g I n T h r A s p A l a V a I T y r l e u P r o I u G I y A l a P r o G I n A I a I l e 377
Db      7017 A T C A T G C T C A A A A A T T A G A T A C C G T T A G C C G C G T A A C G T G A C G T G C T G C T -- 6961
Qy      378  A s p T h r G I n G I u A l a P r o I y S t h r P r o A l a P r o A l a 389
Db      6960 -----G C T T C C G C T C C C G A G C T T C G C C C G C A 6931
RESULT 8
AAL46493
ID      AAL46493 standard; DNA; 691 BP.
AC      AAL46493;
XX
XX      19-JUL-2002 (first entry)
XX
XX      M catarrhalis MCA100170 gene seq ID NO: 1.
XX
XX      Moraxella; vaccine; respiratory tract infection; antiinflammatory;
XX      auditor; antibacterial; otitis media; sinusitis; pneumonia; gene; ds.
XX      Moraxella catarrhalis.
XX
XX      MO200218595-A2.
XX
XX      07-MAR-2002.
XX
XX      28-AUG-2001; 2001MO-CA01221.
XX
XX      28-AUG-2000; 2000US-228294P.
XX      28-AUG-2000; 2000US-228296P.
XX      28-AUG-2000; 2000US-228296P.
XX      29-AUG-2000; 2000US-228438P.
XX      29-AUG-2000; 2000US-228439P.

```

```

PR      29-AUG-2000; 2000US-228440P.
PR      29-AUG-2000; 2000US-228441P.
PR      29-AUG-2000; 2000US-228442P.
PR      29-AUG-2000; 2000US-228443P.
PR      29-AUG-2000; 2000US-228443P.
PR      29-AUG-2000; 2000US-228511P.
PR      29-AUG-2000; 2000US-228512P.
PR      29-AUG-2000; 2000US-228742P.
PR      29-AUG-2000; 2000US-228743P.
PR      01-SEP-2000; 2000US-229465P.
PR      01-SEP-2000; 2000US-229474P.
PR      01-SEP-2000; 2000US-229475P.
PR      01-SEP-2000; 2000US-229478P.
PR      05-SEP-2000; 2000US-229740P.
PR      05-SEP-2000; 2000US-229803P.
PR      05-SEP-2000; 2000US-229804P.
PR      05-SEP-2000; 2000US-229805P.
PR      05-SEP-2000; 2000US-229806P.
PR      05-SEP-2000; 2000US-229809P.
PR      05-SEP-2000; 2000US-229811P.
PR      06-SEP-2000; 2000US-230214P.
PR      06-SEP-2000; 2000US-230250P.
PR      06-SEP-2000; 2000US-230252P.
XX
XX      (AVET ) AVENTIS PASTEUR LTD.
XX
XX      Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;
XX
XX      WPI; 2002-401721/43.
XX
XX      P-P-SDB; AAO17561.
XX
XX      Moraxella polypeptide and polynucleotides useful as vaccine for
XX      immunizing a host e.g. humans against disease e.g. otitis media,
XX      pneumonia, caused by infection of the bacteria
XX
XX      Claim 2; Fig 1; 277pp; English.
XX
XX      The present invention provides the protein and coding sequences of
XX      CC proteins from Moraxella catarrhalis. These can be used to produce
XX      CC vaccines which protect against M. catarrhalis infection, which can cause
XX      CC otitis media, respiratory infection, sinusitis, and pneumonia. The
XX      CC present sequence is a coding sequence of the invention.
XX
XX      SQ Sequence 691 BP; 184 A; 160 C; 161 G; 186 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 4,09e-84 Length: 691
XX      Score: 1034.00 Matches: 197
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 39,384 Indels: 0
XX      DB: 24 Gaps: 0
XX
XX      US-10-088-045-2 (1-502) x AAL46493 (1-691)
Qy      1 MetSerLySPProThrLeuIleYsThrThrLeuIleCyAlaLeuSerAlaLeuMetLeu 20
Db      101 A N C T C A A C C T A C T T G A T A A A A C A C C T A A T T T G C C T T A A G T C A T T A G C T C 160
Qy      21 SerGlyCySerAsnGlnAlaAspIySaIaIaGlnProIySserSerThrVaIaPaIa 40
Db      161 A G T G T T T A G C A A T C A A C G G A A G C C G C C A A A A A G C A G C A G T A G A C G C T 220
Qy      41 A l a I a l e T h r A l a S n A l a P h e a n A l a I a S e r G I n G I u H I s g I n G I u L e u P r o 60
Db      221 G C G C C A A A C A G A A A T G A G A T A T G T G C T C A C A A G A A C A T C A A G C G A G C T G C C T 280
Qy      61 V a l l i e A s p A l a I l e V a l T h r I S a l a P r o G I u V a I P r o P r o V a l A s p A r g A e P h I S 80
Db      281 G T C A T T G A T G C C A T T G T T Y C G C A T G C A C C A A G T T C C A C C A C T T T A C C G T A C C A C 340
Qy      81 P r o A l a l y S a I V a I V a I y S e r G I u T h r V a I g I u I y S a I M e t A r g L e u A l a S p G I y 100
Db      341 C C G C A A A G T G T G T A A A A T G G A A A C G T T G A A A A A G T A T G A T G C T G C G A G A T G C G 400

```

QY 101 ValGluTyrGlnPheThrPheGlyGlyGlnValProGlyGlnMetIleArgValArg 120
DB 401 GTGGAATATCATGTTTGGACATTTGGCGGTCAAGTTCACAGGCGAGATATCGTGCGT 460
QY 121 GluGluAspThrIleGluValGlnPheSerAsnHisProAspSerIleMetProHisAsn 140
DB 461 GAAGCGCACCATCTGAAGTGCAGTTCMAACCCACGATTCMAAATATCCCATATAT 520
QY 141 ValAspPheHisAlaAlaThrGlyProGlyGlyGlnValAlaSerPheThrAlaPro 160
DB 521 GTTACCTTTCACGCTGCACAGGCGCTGGCGGCGGAGAACGCTCATTTACGCGACG 580
QY 161 GlyHisThrSerThrPheSerPheValAlaLeuGlnProGlyLeuTyrValTyrHisCys 180
DB 581 GGTCAATACATCACTTATAGTTTAAAGCTTACAGCTGGTTGTATGCTATCACTGCT 640
QY 181 AlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIle 197
DB 641 GCGGTTCCTCGTTGGCATGCACATTCCTAATGGCATGATATGTTGATT 691

RESULT 9
AAS59518
ID AAS59518 standard; DNA; 37337 BP.
XX
AC AAS59518;
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein encoding DNA #13.
XX
KW SARHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant; ds.
XX
OS Propionibacterium acnes.
XX
PN W0200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Claim 1; SEQ ID NO 13; 1069pp; English.
XX
CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by P. acnes. The disorders include
CC SARHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
CC in infections of bone, joints and the central nervous system, however it
CC is particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acnes in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production

CC of antibodies specific for P. acnes proteins. These antibodies can be
CC used to downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AAU42549-42961, AAU67480 and AAU67481.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 37337 BP; 7213 A; 12017 C; 11261 G; 6843 T; 3 other;

Alignment Scores:
Pred. No.: 5,64e-20 Length: 37337
Score: 349.00 Matches: 92
Percent Similarity: 47.10% Conservative: 46
Best Local Similarity: 31.40% Mismatches: 121
Query Match: 13.29% Indels: 34
DB: 23 Gaps: 10

US-10-088-045-2 (1-502) x AAS59518 (1-37337)

QY 72 ValProProProVal-----AspArgAspHisProAlaValValValMetGlu 89
DB 17911 GTGCTACACACACCTGATGACCGGACGCTTACCGGATGACGCTCCGCGG----- 17964
QY 90 ThrValGluValMetArgLeuAlaAspGlyValGluTyrGlnIleThrPheGly 109
DB 17965 -----GAGTCGGGTGACGAGATGCGCCCGAAACACCATCGACCGGATGACTACAC 18018
QY 110 GlyGlnValProGlyGlnMetIleArgValArgValArgGlyAspThrIleGluValGlnPhe 129
DB 18019 GCGCGTTACATGCGTCCGGTTATTCACGCGCATCGGAGACAGATGAGGTCATCTG 18078
QY 130 SerAsnHisProAspSerIleMetProHisAsnValAspPheHisAlaIleThrGlyPro 149
DB 18079 GTCACACAG-----GCACTAATGAGGTCACTTTGGACTTTCACCGCGTACT----- 18126
QY 150 GlyGlyGlyValAlaGluAlaSerPheThr-----AlaProGlyHisThr 163
DB 18127 -----GTCCTCCCAACAGGCGTATCGCACCATCGCTCCGCGGACAGAG 18171
QY 164 SerThrPheSerPheValAlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAla 183
DB 18172 TTGAGCTACACATTCACCTCGATCGCGCGGATGCTGCTACCATGCTCCACCGCG 18231
QY 184 ProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleValGluProIleGlu 203
DB 18232 CCAATGTCGCGCCCATATTCACAGGACATGTTCCGTCGTCATCGTCCGCGCACAGC 18291
QY 204 GlyLeuProIleValAlaAspGlyGluTyrValMetGlnGlyAspPheThrIleGly 223
DB 18292 ---CTGCTCGGCTATCGAGGATTTTACCTGAGTGGAGACTTACCTCACTGAG 18348
QY 224 LysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla 243
DB 18349 CACAATGGGGCGGAA-----GTTAATATCGCAAGATTCGCAACAGACCCCT 18396
QY 244 GluTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAla 263
DB 18397 GATCTGACGATGTTTCAACGCTCACCGCAACCAAGTACGCTTTGAG---CCCTCAAAAGCC 18453
QY 264 LysValGlyGlnThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSer 283
DB 18454 CGGCTGGAGGAGGAGTGGCATCTGGGTCTCGCGCTGCTCAAGCGGGGCTTCTC 18513
QY 284 PheHisValIleGlyGluIlePheAspLysValHisPheGlnGly----- 298
DB 18514 TTTCACTGCTGGGAGAACCAAGTTTCGACACTGTTTAAAGAGGGCGCCCTAACCCCTTAA 18573
QY 299 ---GlyLysGlyGluAsnHisAsnIleGlnThrIleThrLeuIleProAlaGlyAlaAla 317
DB 18574 CGTGCATATCCGAAGGTGGCGGCTGCGCAAGCCCTGCACTGCGCAGCGCCAGGGCGGT 18633

QY 318 ILeThrGluphelysValAspValProGlyAspTyrValIleuValAspHisAlaIlePhe 337
|||
Db 18634 TTGTGCGAGATGCTTTGAGAGCCGACGACGCTTGTAACTTC--TTc 18690
338 ArgAlaPheAsnlysglyAlaLeuGlyIleuLeuVal 350
:::|||||
Db 18691 GTGAGATGAGCGCTGAAGATTCATTGAGGTG 18729
:::|||||
RESULT 10
AAZ53036
ID AAZ53036 standard; DNA; 399 BP.
XX
AC AAZ53036;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 007 partial DNA sequence SEQ ID NO:43.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX
KW antibacterial; gene therapy; ds.
XX
OS Neisseria gonorrhoeae.
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99MO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
XX
PR 31-JUL-1998; 98US-0094869.
XX
PR 02-SEP-1998; 98US-0098994.
XX
PR 02-SEP-1998; 98US-0099062.
XX
PR 09-OCT-1998; 98US-0103749.
XX
PR 09-OCT-1998; 98US-0103796.
XX
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Frazer C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
XX
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX
PI Tettelin H, Venter JC.
XX
DR WPI; 2000-062150/05.
XX
DR P-PSDB; AAY74274.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
XX
PT vaccines and diagnostics -
XX
PS Claim 7; Page 180; 1453pp; English.
XX
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
XX
CC PCR primers used in the exemplification of the present invention. The
XX
CC polypeptides, the polynucleotides, antibodies and compositions of
XX
CC the invention can be used as vaccines, as diagnostic reagents, and as
XX
CC immunogenic compositions. The polypeptides can be used in the
XX
CC manufacture of medicaments for treating or preventing infection due to
XX
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX
CC presence of Neisseria bacteria, or to raise antibodies. They may also
XX
CC be used to screen for agonists or antagonists, which may themselves
XX
CC have use as antibacterial agents. The polynucleotides of the invention
XX
CC may also be used in gene therapy protocols.
XX
SQ Sequence 399 BP; 118 A; 122 C; 93 G; 66 T; 0 other;
Alignment Scores: 1.64e-12 Length: 399
Pred. No.:

Score: 237.00 Matches: 50
Percent Similarity: 60.58% Conservative: 13
Best Local Similarity: 48.08% Mismatches: 39
Query Match: 9.03% Indels: 2
DB: 21 Gaps: 2
US-10-088-045-2 (1-502) x AAZ53036 (1-399)
QY 398 GlyLeuAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyVal 417
|||
Db 88 GGGCAAAAAGTGTGACATTCATCTGCTGCTGCAAGCAAGGGAAGGCGCC 147
418 ProAsnAlaPheProPheLeuAlaAsnSerAspTyr--LeuAsnAlaAspHisAlaArg 436
|||
Db 148 GGCACGCGCTTCCCTCCGCTTCCGCTGCACTATATATGAACAACCGCACGCTCG 207
437 AlaAlaSerTlleValAlaAsnGlyLeuSerGlyValIleThrValAsnGlyAsnGlnTyr 456
|||
Db 208 CTGCACGACATGCTC--AAAGGACATCAACGATCAATCAAGTCAACGCAAAACCTAC 264
457 GluSerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyr 476
|||
Db 265 AACGATTGATGCCCGCAACCGCATCGCATGCGGACATTGCCCGCTCGCACTTAT 324
477 ThrLeuAsnSerPheGlyAsnLygIlyGlnLeuSerAlaAspValAlaValAla 496
|||
Db 325 ATCATGACACGCTTTCATCAACGCGGCGGAGCGTTACCGAAAGACGTAAACAGGCA 384
497 LysLysThrLys 500
|||
QY 385 AAAGCCAAAAA 396
|||
RESULT 11
ABZ39100
ID ABZ39100 standard; DNA; 552 BP.
XX
AC ABZ39100;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae nucleotide sequence SEQ ID 2789.
XX
KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
KM Neisseria gonorrhoeae.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-1B02069.
XX
PR 12-FEB-2001; 2001GB-0003424.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Fontana MR, Piza M, Maignani V, Monaci E;
XX
PI WPI; 2003-058415/05.
XX
DR P-PSDB; ABP78130.
XX
PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX
PT medicament for treating or preventing N. gonorrhoeae infection -
XX
PS Disclosure; Page 392; 815pp; English.
XX
XX
CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX
CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX
CC antibodies that specifically bind to the proteins. The composition
XX
CC comprising the protein, nucleic acid or antibody is useful for the
XX
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
XX
CC infection, this may be in the form of a vaccine or gene therapy.
XX
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid

CC molecules of the invention.

SQ Sequence 552 BP; 175 A; 167 C; 116 G; 94 T; 0 other;

Alignment Scores:

Pred. No.:	2,536-12	Length:	552
Score:	237.00	Matches:	50
Percent Similarity:	60.56%	Conservative:	13
Best Local Similarity:	48.08%	Mismatches:	39
Query Match:	9.03%	Indels:	2
	25	Gaps:	2

US-10-088-045-2 (1-502) x AB239100 (1-552)

```

QY 398 GilybalaThrTyraSpSerAnCybAlaLaCySHiGInProAspGlyLySgLYVal 417
    |||:::|||||
DB 175 GGGCAAAAAGTACGAATCCACTCCGCTCCACGGCAAAAGGGGAAGGGCGC 234
QY 418 ProAnaAlaPheProProLeuAlaAnSerAspTyR--LeuAnaAlaSPHisAlaArg 436
    |||:::|||||
DB 235 GGCACCTGGCTTCTCCGCTTTCCGTCGACCTATATATGAACAACCGACGCTCTG 294
QY 437 AlaAlaSerIleValAlaAnGlyLeuSerGlyLySgIleThrValaAnGlyLySgInTyR 456
    |||:::|||||
DB 295 CTGCACACATGTC--AAAGCATCAACGCTACAAATCAAGTCAACGGCAAAACCTAC 351
QY 457 GluSerValMetProAlaIleAlaLeuSerAspGInGInIleAlaAnValIleThrTyR 476
    |||:::|||||
DB 352 AACGATTCATGCGCGCAACCGCCATCCGATCGGACATTCGCGCGCTGCCACTTAT 411
QY 477 ThrLeuAnSerPheGlyAnuLySgIlyGInLeuSerAlaAspAspValAlaLySAla 496
    |||:::|||||
DB 412 ATCATGAACGCTTTGACACACGCGCGGAGCGTACGAAAGAAAGCTAAACAGGCA 471
QY 497 LySlyThrLyS 500
    |||
DB 472 AAAGCAAAAA 483

RESULT 12
AAZ53038
ID AAZ53038 standard; DNA; 396 BP.
XX
AC AAZ53038;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:47.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;

```

PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tectelin H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR P-RSDb; AAY74276.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
PS Claim 7; Page 180-181; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

SQ Sequence 396 BP; 118 A; 125 C; 92 G; 61 T; 0 other;

Alignment Scores:

Pred. No.:	3,666-11	Length:	396
Score:	222.00	Matches:	44
Percent Similarity:	56.31%	Conservative:	14
Best Local Similarity:	42.72%	Mismatches:	45
Query Match:	8.45%	Indels:	0
	21	Gaps:	0

US-10-088-045-2 (1-502) x AAZ53038 (1-396)

```

QY 398 GilybalaThrTyraSpSerAnCybAlaLaCySHiGInProAspGlyLySgLYVal 417
    |||:::|||||
DB 88 GGGCAAAAAGTACGAATCCACTCCGCTCCACGGCAAAAGGGGAAGGGCGC 147
QY 418 ProAnaAlaPheProProLeuAlaAnSerAspTyRLeuAnaAlaSPHisAlaArgAla 437
    |||:::|||||
DB 148 GGAACCATGTTTCCGCGCTCTACCGCTCCGATTCATCATGAAACCGCAGTGCTG 207
QY 438 AlaSerIleValAlaAnGlyLeuSerGlyLySgIleThrValaAnGlyLySgInTyRGu 457
    |||:::|||||
DB 208 CTGCACACATGTCGTCAAAGCATCAACGATCAAGTCAACGCAAAACCTACAC 267
QY 458 SerValMetProAlaIleAlaLeuSerAspGInGInIleAlaAnValIleThrTyRThr 477
    |||:::|||||
DB 268 GGAATTCATGCGCGCCACCTGCATCAGCATGCGAGCATTCGCCCGCTGCCACTTATAC 327
QY 478 LeuAnSerPheGlyAnuLySgIlyGInLeuSerAlaAspAspValAlaLySAlaLyS 497
    |||:::|||||
DB 328 ATGAACGCTTTGACACACGCGCGGAGCGTACCGAAAGAAAGCTAAACAGGCAAAA 387
QY 498 LySlyThrLyS 500
    |||
DB 388 AACCAAAAA 396

RESULT 13
AAZ53037
ID AAZ53037 standard; DNA; 402 BP.
XX
AC AAZ53037;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:45.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

```


KM antibacterial; gene therapy; ds.
 XX Neisseria meningitidis.
 OS
 XX MO9957280-A2.
 PN
 XX 11-NOV-1999.
 PD
 XX 30-APR-1999; 99WO-US09346.
 PF
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 04-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarfelli M;
 PI Tettelin H, Venter JC;
 DR WPI; 2000-062150/05.
 DR P-PSDB; AAY74275.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS
 XX Claim 7; Page 180; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 XX
 XX Sequence 402 BP; 122 A; 125 C; 93 G; 62 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 3 73e-11 Length: 402
 Score: 222.00 Matches: 44
 Percent Similarity: 56.31% Conservative: 14
 Best Local Similarity: 42.72% Mismatches: 45
 Query Match: 8.45% Indels: 0
 DB: 21 Gaps: 0
 US-10-088-045-2 (1-502) x AA253017 (1-402)
 QY 398 Gtlysalathtyrspseranrcyvalaalacyahieglnproaspilyysglyval 417
 DB 88 GGGCAAAAGTGAACATCAACTCGCTGCTGCGCAAGGCAAGGCGCC 147
 QY 418 ProasnaAlpheproProleuAlaasnsrAspTYrleuAsnAlaAspHisAlaArgAla 437
 DB 148 GGAACCATGTTCCGCGCTCAACCGCTCGAATTCAATCATGAAAAAACCAGAGTCTG 207
 QY 436 AlaSerIleValAlaAsnGlyLeuSerGlyValIleThValAsnGlyAsnGlnTYrGlu 457
 DB 208 CTGCACGATGTCAAAGGCAATCAACGGGACATCAATAAAGTCAACGGCAAACTTCAAC 267
 QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTYrThr 477

DB 268 GGATTCATGCGCCGCAACCGCATTCAGCGCATTCGCCGCTGCCCTATATAC 327
 QY 478 LeuAsnSerPheGlyAsnTySGlyGlnLeuSerAlaAspAspValAlaAsnAlaTyS 497
 DB 328 ATGAACGCTTTGACACACGCGCGGAGAGCTTACGAAAAAGACGTAAACAGCGCAAA 387
 QY 498 LysThrTyS 500
 DB 388 AGCAAAAAA 396
 RESULT 14
 ID AAA81733/C
 ID AAA81733 strand; DNA; 12893 BP.
 AC AAA81733;
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_280 SEQ ID NO:280.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO20002430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US23573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 PI Frazer CM, Hickey E, Petersen J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 DR WPI; 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 PS
 XX Claim 7; Page 1561-1565; 1760pp; English.
 XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and

CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

SO Sequence 12893 BP; 2829 A; 3164 C; 3741 G; 3159 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	3,91e-09	12893	64	32	84	42
Percent Similarity:	222.00					
Best Local Similarity:	43.44%					
Query Match:	28.96%					
DB:	21					6

US-10-088-045-2 (1-502) x AAA81733 (1-12893)

```

QY 301 G|YGLUASNH|ASNI|EG|N|Th|Leu|le|Pro|Ala|G|Y|Ala|Ala|le|Th|G|U 320
DB 10503 GGAACAACACATCAAGTCAAA-----CCTTATAGTGATTAATTAAACGAG 10456
QY 321 Phe|y|Val|Asp|Val|Pro|G|y|Asp|Tyr|----- 329
DB 10455 TACGGCTTCGCTCGCTTG--CGCTACTATTGTGCTGCGGCTTCGCTCGCTTATC 10397
QY 330 Val|Leu|Val|Asp|His|Ala|le|Phe|Arg|Ala|Phe|Asn|y|S|G|Y|Ala|Leu|G|Y|le|Leu|Ys 349
DB 10396 CTGATTTTGTATCACTACTACTACTGATTCGGCAAAAATCTTCAACCGCTACGC 10337
QY 350 Val|G|U|G|Y|G|U|G|U|A|N|H|S|G|U|I|e|T|y|S|e|r|H|S|I|S|G|I|N|T|H|-----Asp|Ala|Val 367
DB 10336 -----AAAATCAACAACAAGTTTCAAAATGAGAAACACATGACACCAACC 10286
QY 368 Tyr|Leu|Pro|G|U|-----G|Y|Ala|P|ro|G|I|N|Ala|le|Asp|Thr 379
DB 10285 CGACTGCGGACCGCCCTCGCTTGGGCTGCTTGGCCGCCGCTTGGCCGCGACAC 10226
QY 380 G|N|G|U|A|l|A|P|ro|y|Th|P|ro|Ala|P|ro|Ala|A|S|N|L|e|U|G|I|N|I|e|Y|S|A|G|Y|Ys 399
DB 10225 AGCATCATGACAAA-----GGGCAA 10205
QY 400 Al|A|Th|T|y|T|A|S|e|r|A|N|C|y|A|Ala|C|y|H|S|G|I|N|P|ro|A|S|G|Y|S|G|Y|Ala|P|ro|A|S|N 419
DB 10204 AAAGTATGACAAATCAACTCGCTCGCTGCGCAAAAAGGCGCAAGCCGCGAACC 10145
QY 420 Al|A|P|he|P|ro|P|ro|Leu|Ala|A|S|e|r|A|P|T|y|L|e|u|A|S|N|Ala|A|S|P|H|S|A|l|A|S|e|r 439
DB 10144 ATGTTTCGCGCTCTACCGCTCGCACTTATCATGAAAAACCGCGAGTCTGCTGAC 10085
QY 440 I|L|e|V|A|l|A|S|N|G|Y|L|e|U|S|e|r|G|Y|Y|S|I|e|T|H|V|A|l|A|S|N|G|Y|S|N|G|I|N|Y|G|U|S|e|r|V|A|l 459
DB 10084 AGCATGATCAAGGATCAACGGTACATCAAGTCAACGCAAAACCTTCAACGGAATTC 10025
QY 460 Me|P|ro|Ala|le|Ala|le|U|S|e|r|A|P|G|I|N|I|e|Ala|A|S|N|V|A|l|le|T|H|T|y|T|H|Leu|A|S|N 479
DB 10024 ATGCGCCGCAACCGCATCAGCGATGCGGACATTCGCCCTCGCACTTATCATGATAC 9965
QY 480 Se|r|P|he|G|Y|A|S|N|y|S|G|Y|G|I|N|Leu|S|e|r|A|A|S|P|H|S|A|l|A|S|e|r|V|A|l|A|l|y|S|I|e|Y|S|Thr 499
DB 9964 GCTTTTACAAACGGCGGAGAGCGTTACCGAAAAAGACGTTAAACAGGCAAAAGCAA 9905
QY 500 Lys 500
DB 9904 AAA 9902

```

RESULT 15

AAAF21608/c standard; DNA; 349980 BP.

AAAF21608;
 13-MAR-2001 (first entry)
 Neisseria meningitidis B nucleotide sequence SEQ ID NO.109.

KM Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KM ds.

OS Neisseria meningitidis.

PN MO200066791-A1.

PD 09-NOV-2000.

PF 08-MAR-2000; 2000MO-US05928.

PR 30-APR-1999; 99US-0132068.

PR 08-OCT-1999; 99MO-US23573.

PR 28-FEB-2000; 2000GB-0004695.

PA (CHIR) CHIRON CORP.

PA (GENO-) INSTR GENOMIC RES.

PI Pizza M, Hickey E, Peterson J, Tettein H, Venter JC, Masignani V;
 PI Galeotti C, Mora M, Ratti G, Scarbelli M, Scariato V, Rappoli R;
 PI Frazer CM, Grandi G;

XX MPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections -

XX Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.

XX Sequence 349980 BP; 82523 A; 82940 C; 96712 G; 87805 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	3,28e-07	349980	64	32	84	42
Percent Similarity:	222.00					
Best Local Similarity:	43.44%					
Query Match:	28.96%					
DB:	21					6

US-10-088-045-2 (1-502) x AAF21608 (1-349980)

```

QY 301 G|YGLUASNH|ASNI|EG|N|Th|Leu|le|Pro|Ala|G|Y|Ala|Ala|le|Th|G|U 320
DB 148400 GGAACAACACATCAAGTCAAA-----CCTTATAGTGATTAATTAAACGAG 148353
QY 321 Phe|y|Val|Asp|Val|Pro|G|y|Asp|Tyr|----- 329
DB 148352 TACGGCTTCGCTCGCTTG--CGCTACTATTGTGCTGCGGCTTCGCTCGCTTATC 148294

```



```
DB 748040 ATGTTTCCGCGCCCTACCGCTCCGACTTCATCATGAAAAAACCAGAGTGTCTGTGCAC 747981
QY 440 ILevalAlaenGlyLeuSerGlySerIleThrValaenGlyAsnGlnTyrGluSerVal 459
DB 747980 AGCATGCGTCAAGGATCAAGCGTACATCAAAAGTCAACGGCAAAACCTTACACGGATTTC 747921
QY 460 MetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrIleuAsn 479
DB 747920 ATGCCCCGCAACCGCATCAGCATGCGGACATTCGCCCGCTCGCCACTTATATCATGAAAC 747861
QY 480 SerPheGlyAsnLeuGlyGlyGlnLeuSerAlaAspValAlaIleAlaIleLysLysThr 499
DB 747860 GCCTTTGACAAACGGCGCGGAGAACGCTTACGAAAAAGACGTAAAAACGCAAAAGCAAA 747801
QY 500 Lys 500
DB 747800 AAA 747798

RESULT 17
AAZ53034
ID AAZ53034 standard; DNA; 341 BP.
AC AAZ53034;
XX
XX
DT 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:39.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
XX Neisseria meningitidis.
OS
XX WO957280-A2.
XX
XX PD 11-NOV-1999.
XX
XX PF 30-APR-1999; 99WO-US09346.
XX
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalzo E, Scarfelli M;
XX PI Tettelein H, Venter JC;
XX
XX DR MPI, 2000-062150/05.
XX DR P-PsDB; AAY74272.
XX
XX PT Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics
XX
XX PS Claim 7, Page 178-179, 1453pp; English.
XX
XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54733 represent
XX CC PCR primers used in the exemplification of the present invention. The
XX CC polypeptides, the polynucleotides, antibodies and compositions of
XX CC the invention can be used as vaccines, as diagnostic reagents, and as
XX CC immunogenic compositions. The polypeptides can be used in the
XX CC manufacture of medicaments for treating or preventing infection due to
```

```
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 341 BP; 91 A; 113 C; 79 G; 58 T; 0 other;
Alignment Scores:
Pred. No.: 8.05e-08 Length: 341
Score: 184.00 Matches: 36
Percent Similarity: 55.95% Conservative: 11
Best Local Similarity: 42.86% Mismatches: 37
Query Match: 7.01% Indels: 0
DB: 21 Gaps: 0

US-10-088-045-2 (1-502) x AAZ53034 (1-341)
QY 398 GlyLeuAlaThrTyrAspSerAsnGlyAlaAlaCysHisGlnProAspGlyLysGlyVal 417
DB 88 GGGCAAAAGTGTACCAATCCAACTGCGCTGCCACGGCAAAAGGGCGAAGCGCGC 147
QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla 437
DB 148 GGAACCATGTTTCGCGCGCTTACCGCTCCGACTTCATCATGAAAAAACCAGAGTGTCTG 207
QY 438 AlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValaenGlyAsnGlnTyrGlu 457
DB 208 CTGCACAGCATGTGCAAGGACATCAACGCTACATCAAAAGTCAACGGCAAAACCTACAAAC 267
QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThr 477
DB 268 GGATTCATGCGCGGACCGCATCAAGCGATCGACATTCGCCCGCTGCCACTTATATATC 327
QY 478 LeuAsnSerPhe 481
DB 328 ATGAACGCTTT 339

RESULT 18
AAZ53035
ID AAZ53035 standard; DNA; 341 BP.
AC AAZ53035;
XX
XX
DT 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 007 partial DNA sequence SEQ ID NO:41.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
XX Neisseria meningitidis.
OS
XX WO957280-A2.
XX
XX PD 11-NOV-1999.
XX
XX PF 30-APR-1999; 99WO-US09346.
XX
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
```

PI Petersen J, Pizsa M, Rapuoli R, Ratti G, Scalato E, Scarselli M,
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR P-PSDB; AAY74273.
XX
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
PS
XX
XX Claim 7; Page 179; 1453pp; English.

AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX
XX SQ Sequence 341 BP; 90 A; 114 C; 79 G; 58 T; 0 other;

Alignment Scores:
Pred. No.: 8,05e-08 Length: 341
Score: 184.00 Matches: 36
Percent Similarity: 55.95% Conservative: 11
Best Local Similarity: 42.86% Mismatches: 37
Query Match: 7.01% Indels: 0
DB: 21 Gaps: 0

US-10-088-045-2 (1-502) x AA253035 (1-341)

QY 398 GILYSAATHrTYrAspSerAsnCysAlaAlaCyHieGlnProAspGlyLyseGlyVal 417
DB 88 GGGCAAAAAGTACGATCAATCCACTCGCTGCGCAAGGCAAAAGGCGCCG 147
QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspTYrLeuAsnAlaAspHisAlaArgAla 437
DB 148 GGAACCATGTTCGCGCGCTACCGCTCGACTCATCATGAAAAACCGCAGGTGCTG 207
QY 438 AlaSerIleValAlaAsnGlyLeuSerGlyIleThrValAsnGlyAsnGlnTYrGlu 457
DB 208 CTGCACAGCATGTCACAAAGCATCAACGGTACATCAAGTCAACCGCAAAACCTACAC 267
QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTYrThr 477
DB 268 GGATTCATGCCCGCCACATCGCATCGGACATTCGCGCGCTCGCCACTTATATC 327
QY 478 LeuAsnSerPhe 481
DB 328 ATGAACGCTTT 339

RESULT 19
ABLS2917
ID ABLS2917 standard; DNA; 4290 BP.
XX
XX ABL52917;
AC
XX
XX 16-JUL-2002 (first entry)
DT
XX
XX 2-keto-D-gluconate dehydrogenase coding sequence.
DE
XX
XX Cell membrane bound; 2-keto-D-gluconate dehydrogenase; enzyme;
KW 2,5-diketo-D-gluconate; gene; ds.
XX
XX Unidentified.
OS
XX
XX Key Location/Qualifiers
FH

FT CDS 337..3991
FT
FT /tag= a
FT /product= "2-keto-D-gluconate dehydrogenase"
FT /note= "Contains one intron"
FT
FT exon
FT 337..900
FT /tag= b
FT /number= 1
FT intron
FT 901..926
FT /tag= c
FT /complice= (5'site:NO,3'site:NO)
FT
FT /number= 1
FT exon
FT 927..3988
FT /tag= d
FT /transl_except= (pos:2580..2581, aa:Met)
FT /number= 2

PN KR2000019366-A.
XX
XX
XX 06-APR-2000.
XX
XX PD
XX
XX PF 10-SEP-1998; 98KR-0037413.
XX
XX PR 10-SEP-1998; 98KR-0037413.
XX
XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
XX
XX PA
XX Shin YC, Bahn JG, Yeom DY;
XX
XX PI
XX
XX WPI; 2001-088033/10.
XX
XX DR P-PSDB; AAM48440.
XX
XX PT Base sequence of novel cell membrane-bound 2-keto-D-gluconate
XX dehydrogenase gene and method for conversion into
XX 2,5-diketo-D-gluconate - Noabstract
XX
XX PS Claim 2; Page 6-8; 16pp; Korean.
XX
XX CC The present invention relates to a novel cell membrane-bound
XX 2-keto-D-gluconate dehydrogenase and a method for conversion into
XX 2,5-diketo-D-gluconate. The present sequence is the coding sequence for
XX the 2-keto-D-gluconate dehydrogenase.

SQ Sequence 4290 BP; 992 A; 1177 C; 1226 G; 895 T; 0 other;

Alignment Scores:
Pred. No.: 1.56e-05 Length: 4290
Score: 175.00 Matches: 109
Percent Similarity: 36.10% Conservative: 52
Best Local Similarity: 24.44% Mismatches: 163
Query Match: 6.66% Indels: 124
DB: 23 Gaps: 20

US-10-088-045-2 (1-502) x ABL52917 (1-4290)

QY 134 AspSerIleMetProHisAsnValAspPheHisAlaIleThrGlyProGlyGly----- 151
DB 2620 GACGTTAAGCGCTCCGCGACAGCAGACCGATCTGCATTTGCGCCGAGGGAACAGGT 2679
QY 152 -----Gly-AlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSe 167
DB 2680 CGCCATGCTTCGATGTCAGGATGCGACACCGCTCCGCGCAGTGCACACCGCTTCG 2739
QY 167 r-----PheGlyAlaLeuGlnProGlyLeuTYrValTYrHisCysAlaValAlaPro-- 184
DB 2740 CGCGGTTATGCGATGCGCTCGCGATGGTGTATTCATTCACCAATATCATCCCTTCG 2799
QY 185 -----ValGlyMetHis-----IleAlaAsnGlyMet 193
DB 2800 TGCCGACGCGATCGGCGCTTACAGGAGCATGTAATTTTCGACGGCGGTACCGCATGGCT 2859
QY 193 tTYrGly-----LeuIleLeuValGlnProGlySerGlyLeuProGlyValAs 209
DB 2860 ACGCGTACGCGCGCGCATCTTATCCGCGATGCTTACACCTTACAGCAAGATCAC 2919

OY	209	plysdlu-----	-----TyrTyrValMetGlnGlyAspPheTyrThLysG1	223
Db	2920	GGATGAAGATTTCATGCGCTTATTA	CTTACTTATTCATTTGATGCGCGTGAAGCCGGTTGACCA	2979
OY	223	YLYeTyrclgylglnGlyLeuGln-	---ProPheAspMetGluYalaIle-----	239
Db	2980	GAANAACCGACAAACACAGCTGCT	CTTCCCTTCCCTTTAAACCGCCTTACACATGTTCTTG	3039
OY	240	-----	-----AcsGluAspAlaGluTyrVal-----	246
Db	3040	GAACCTGATGTTGGCGAGACGATTA	ACCCTTACCTTGACCGATACAGCCACGAGTCTGATG	3099
OY	247	-----	-----ValPheAsnGlySerVal-----	252
Db	3100	GAATCGGGATTAATCTATCTGTAA	CGGCTTGCCGACATGTAACACCTGATACCGCGG	3159
OY	253	-GlyAlaLeuThrGlyGluAsnAla	IleuYalaIlyValGlyIuThrValArgLeuPh	272
Db	3160	CGGATATTATGATGACGAGGACG	AGCAGCCGCGCTGCTGGTGGCCGATGGCAGCTG	3219
OY	272	evalGlyangGlyGlyProAsnLeu	ThrSerSerPheHisValIleGlyGluIlePheAs	292
Db	3220	GTAC-----	-----GGCCTAATCATCATCC-----	GA 3243
OY	292	plyeValHisPheGlnGlyGlyLy	SerGlyGluAsnHisAsnIleGlnThrThLeuIlePr	312
Db	3244	TGCGATTAGCGGATCGGTGGCTG	CGCTTAACATAGCTGTTCAAT--ACCTTAAGAC	3301
OY	312	calaglyGlyAlaIleThnGluPh	elyValAspAlaProGlyAspTyrValLeuVa	332
Db	3302	GGCCGGG-----	-----	3308
OY	332	lAspHisAlaIlePheArgAlaPhe	AsnLySerGlyAlaLeuGlyIleLeuLyS--ValG1	351
Db	3309	-----	-----CTAGAGGCGAA--AACAGCGCGCGGCGGCGGACGCGAAAGCGGTGA	3354
OY	351	uGlyGluGluAsnHisGluIleTyr	SerHisIlySerGlnThrAspAlaValIlyLeuProG1	371
Db	3355	GCATAGCTGACGATCTCTGTCG	ACGACGAGATTGAACGATAGCGATTTCCTTAAAG	3414
OY	371	uGlyAlaPro-----	-----GlnAlaIleAspThrGlnGluLapProLySerThrProAlaProAl	389
Db	3415	CACAGACGCCAATCCGTGACG	AGAGGACACGCCAACCCGCTTACAGCTTTGGCAACACGC	3474
OY	389	AsnLeuGlnGlnGlnIleYala-	-----	397
Db	3475	GGATTGGAGAAACACATACCG	CGGACGTAACGCCAATACCTACATGACGA	3534
OY	398	-GlyYValAlaThrTyrAspSer	AsnCyAlaAlaGlyHisGlnProAspGlyLySgLyVa	417
Db	3535	TGGCCCGGACCTGTTCAGCGG	CAATCGGCCACGCTCATACGCCAGCGAGCCGGGAG	3594
OY	417	lProAsn--AlaPhePro	LeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaIar	436
Db	3595	CGCTAATCAGGCTTACCTTCG	CTGTTCACAAATAC--GGACGGGCAATGCAATACCC	3651
OY	436	GAlaAlaSerIleValAla-	-----AsnGlyLeuSerGlyLyS1	449
Db	3652	AGCCATCTGATTGGCGCTATTCT	TTGGCGTCGACGCAACCCCGCGGCGTCAAGT	3711
OY	449	eThrValAsnGlyAsnGlnTyr	GluSerValMetProAlaIleAlaLeuSerAspGlnG1	469
Db	3712	ACTGATGCCGGC-----	-----TTAGCTACCCCTTATGTTGACAACTGAGAGTCCCA	3765
OY	469	nIleAlaAsnValIleThrTyr	ThrLeuAsnSerPheGlyAsnLySgIyGlnLeuSe	489
Db	3766	GGTGGCTGATATCACTAAATTG	TGCTGCTGGCGCATTAACGATATCAGAGGTACCGGTAG	3825
OY	489	rAlaAspAspValAla	494	
Db	3826	TGCAGGTGACGTTGCC	3841	

RESULT	20
AAZ53033	
ID	AAZ53033 standard; DNA; 342 BP.
XX	
AC	AAZ53033;
XX	
DT	21-MAR-2000 (first entry)
XX	
DE	Neisseria gonorrhoea ORF 007 partial DNA sequence SEQ ID NO:37.
XX	
KW	Neisseria meningitidis; Neisseria gonorrhoea; antigen; vaccine;
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW	antibacterial; gene therapy; de.
OS	Neisseria gonorrhoea.
XX	
PN	W09957280-A2.
XX	
PD	11-NOV-1999.
XX	
PF	30-APR-1999; 99NC-US09346.
XX	
PR	01-MAY-1998; 98US-0083758.
PR	31-JUL-1998; 98US-0094869.
PR	02-SEP-1998; 98US-0098994.
PR	02-SEP-1998; 98US-0099062.
PR	09-OCT-1998; 98US-0103749.
PR	09-OCT-1998; 98US-0103794.
PR	09-OCT-1998; 98US-0103796.
PR	25-FEB-1999; 99US-0121528.
XX	
PA	(CHIR) CHIRON CORP.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M,
PI	Petersen J, Pizzo M, Rappoldi R, Ratti G, Scatato E, Scarselli M,
PI	Tetzelin H, Venter JC;
XX	
DR	WP1; 2000-062150/05.
XX	
PT	P-PSDB; AAY74271.
PT	Novel Neisserial polypeptides predicted to be useful antigens for
PT	vaccines and diagnostics -
XX	
ES	Claim 7; Page 178; 1453pp; English.
XX	
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC	represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54473 represent
CC	PCR primers used in the exemplification of the present invention. The
CC	polypeptides, the polynucleotides, antibodies and compositions of
CC	the invention can be used as vaccines, as diagnostic reagents, and as
CC	immunogenic compositions. The polypeptides can be used in the
CC	manufacture of medicaments for treating or preventing infection due to
CC	Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC	presence of Neisseria bacteria, or to raise antibodies. They may also
CC	be used to screen for agonists or antagonists, which may themselves
CC	have use as antibacterial agents. The polynucleotides of the invention
CC	may also be used in gene therapy protocols.
XX	
SQ	Sequence 342 BP; 85 A; 108 C; 85 G; 64 T; 0 other;
Alignment Scores:	
Pred. NO.:	9.78e-07 Length: 342
Score:	172.00 Matches: 39
Percent Similarity:	58.14% Conservative: 11
Best Local Similarity:	45.35% Mismatches: 33
Query Match:	6.55% Indels: 3
DB:	21 Gaps: 2
US-10-088-045-2 (1-502) X AAZ53033 (1-342)	
398 GlyAlaThrThrArgSerSerCysAlaAlaCysHISGlnProArgGlyLysGlyVal 417	

```
Db      88 GGGCAAAAGTGCATGCACTGCATGCGCCGCAAGAAAGGAGGCGC 147
      418 ProAspAlaIleProPheLeuAlaIleSerAsp---TyrLeuAsnAlaIlePheAlaArg 436
      148 GGCACGCGCTTCTCCGCTTTCCGGCTGGACTGATTATGAAACCAACCGCAGCTCCTG 207
      437 AlaAlaSerIleValAlaIleAsnGlyLeuSerGlyLysIleThrVal-AsnGlyAsnGlnTy 456
      208 CTCACAGCATGTGTC---AAGGCATCGACGCGACATTCAAAGTGGAGCGGCAAAACCTA 264
      456 rGluSerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTy 476
      265 CGACGATTATTCGCCGCAACCCGACATGCGGATGCGGACATTCGCCGCTCCGCACTTA 324
      476 rThrLeuAsnSerPhe 481
      325 TATCATGAACGCTTT 340
```

RESULT 21

AAQ36687

ID AAQ36687 standard; DNA; 1437 BP.

XX AAQ36687;

DT 25-MAR-2003 (updated)

DT 08-JUN-1993 (first entry)

XX Sequence encoding cytochrome c-553 (CO).

XX Cytochrome; oxidative fermentation; electron transfer system; ss.

XX Gluconobacter suboxydans IFO 12528.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

SQ Sequence 1437 BP; 291 A; 466 C; 437 G; 243 T; 0 other;

Alignment Scores:

Pred. No.:	3,53e-05	Length:	1437
Score:	164.00	Matches:	109
Percent Similarity:	35.688	Conservative:	48
Best Local Similarity:	24.77%	Mismatches:	166
Query Match:	6.25%	Indels:	117
DB:	14	Gaps:	23

US-10-088-045-2 (1-502) x AAQ36687 (1-1437)

```
QY 149 ProGlyGlyGlyAlaGluAlaSerPheThrAlaProGlyHis-----ThrSerThrPhe 166
      75 CCTCATGCGCGGTCTTTCGCGCGCCGATGCCGAGACGCGGATGAATGATGATCA 134
      167 SerPheLeuAlaLeuGlnProGlyLeuTyThrValTyThrHis-CysAlaValAlaProValG 186
      135 GCGTGGCGAATATGTGTCGCCGCTGTGGACTGTCATGCGCTGCATACGCGCTTCACGG 194
      186 yMetHisIleAlaAsnGly-----MetTyGlyLeu-- 196
      195 ACAGCCCTATGCGGCGGTCTGAGATCAAGACCCGATCGGACGATGATTCAGCA 254
      197 -IleLeuValGluProGlyGluGlyLeuProGlyValAspGlyGlyTyThrValMetG 216
      255 CATCACGCGGACCCGAAACAGGATATCGCAAC-----TACACCTGGA 299
      216 nGlyAspPheTyThrTyGlyGlyLysTyGlyGluGlnGlyLeuGlnProPheAspMetG 236
      300 GGAC-----TTCACCAAGGCT-----CTCG 320
      236 yIysAlaIleArgGluAsp----- 242
      321 CAAGGCAATCCGAAAGATGTCGCAAGGCTATCCGCGCATCCATCTGAATTCG 380
      243 -----AlaGluTyThrValPheAsnGlySerValGlyAl 254
      381 AGCTGTCCGATGACGACATCCGTCGATGATGCTTCTTCATCAGCGGCGTGAAGC 440
      254 aleuThrGlyGluAsn-----AlaLeuValAlaIysVal 265
      441 GGTGCGCTGCAAGAACGCGCCGACATCAGCTGCGCTTCATGCGCTGCGCGCT 500
      265 lGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheH 285
      501 GGGCATGTGCGCGCGCATGTCCTCCGAGCATGACGCGCGGTGACAAAGCATCAG 560
      285 s-----ValIleGlyGluLeuPheAspLysValHisPheGluGlyGlyLysG 301
      561 TGATCCGAAAGTTCGCGCGCGCATCTC-----GTCAACGCGCGCGCATTCGCG 614
      301 yGluAsnHisAsnIleGln-----Th 308
      615 CGAATGCCACACGCGCGCGGTTTCGGAATGAGGTCAAGGCTTACGCAACGCGTGGCG 674
      308 rThrLeuIleProAlaGlyGlyAlaAlaIleThrGluPheLysValAspValProGlyAs 328
      675 CAACGCTATCTGCGCGGTGTGTCACGATGCAACTGG-----ATGCAACGAGCCT 728
      328 rTyThrValLeuValAsnHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLe 348
      729 GCGGCAACACAGGACGACGCGGTGCGTGTGATGAATGATGATGATGATGATGATGATG 788
      348 yIysValGluGlyGluGluAsnHisGln-----I 358
      789 CAAG---ACGCGCGGTATGACCACTCCGCTTCGTTGCGCATGCGGACGCGTGGCG 845
      358 eTyThr-----HisGlyGlnThrAsp-----AlaValTyThr----- 369
      846 CTACGACGACGACGATGTCGACGACGATTCGCGCGGACGACGACGATTCCTCAAGAG 905
      370 -----ProGluGlyAlaProGluAlaIleAspThrGlnGluAlaProGlyThr 385
```



```

Db      906 CATCCCGCCGTCGCGAAGCAAGAACTTCGT-----CAGATGACGGTCAGAC 956
Qy      385 rProAlaProAlaLeuInglInleValAlaGlyValAlaThrTyraSerArg 405
Db      957 GACCGCTCTGTGTAACAAGGTGGCCAGGCGCAAGGTGGCAAGTCTATTCGACAA 1016
Qy      405 nCyAlaAlaCyAlaInglInleValAlaGlyValAlaThrTyraSerArg 425
Db      1017 CTGGCGCATCTGCGACATGAACGATGATAGGGGCTCAACCGCATGTTCCCGCTGGC 1076
Qy      425 a---AAsSerAspTyrlLeuAlaAspHlaAlaArgAlaAlaSerlleValAlaAngl 444
Db      1077 AGCGAACCCGGTCCTCATTCACGACATCCGACCTGGCTCGCATATGCTAGCTTCGG 1136
Qy      444 yLeuSerGlyLysIleThrValAlaAnglValAnglInTyrgluserVal---MetProAlaIle 463
Db      1137 T-----GGCATCTGCGCCCGACCAACAGCGACCGTCCGCTTCCATGCCCGGGTT 1190
Qy      463 eAlaA-----LeuSerAspGlnGlnIleAlaAsnValIleThrTyrlLeuAsnSerPh 481
Db      1191 CAAAGATCACCCTCTCCGACCGAGATGGCGATGTCGTAACCTTTATGGCGAAGGCTG 1250
Qy      481 eGlyAsnIleGly---GlyGlnLeuSerAlaAspAspValAlaLysAlaLysThr 499
Db      1251 GGGCAACAGGCAACCGGGAACCGTCTCGCTCCGATATCCAGAACTGCGCACAGCG 1308

```

RESULT 22

AAQ48234 standard; DNA; 1437 BP.

```

ID      AAQ48234 standard; DNA; 1437 BP.
AC      AAQ48234;
DT      13-JAN-1994 (first entry)
XX
XX      Cytochrome C gene.
DE
XX      C-553; improved produ.; production; oxidative fermentation; ss.
XX
XX      Gluconobacter suboxydans IFO 12528.
XX
XX      Key      Location/Qualifiers
FH      sig_peptide      1..108
FT      mat_peptide      109..1437
FT      /*tag= a
FT      /*tag= b
XX
XX      JP05049480-A.
XX
XX      02-MAR-1993.
XX
XX      14-JUN-1991; 91JP-0238579.
XX
XX      14-JUN-1990; 90JP-0154096.
XX
XX      (ASAH ) ASAH CHEM IND CO LTD.
XX
XX      WPI; 1993-169634/21.
XX
XX      P-PSDB; AAR39355.
XX
XX      Cytochrome C gene - comprises specified sequence of 1473 bases
XX      PT which is inserted in host cells and is used to produce cytochrome C
XX      PT by oxidative fermentation
XX
XX      Claim 1; Page 13-15; 20pp; Japanese.
XX
XX      The sequence is that of a cytochrome C gene from Gluconobacter
XX      CC suboxydans IFO 12528 which may be used to give improved produ. of
XX      CC cytochrome C by oxidative fermentation.
XX
XX      Sequence 1437 BP; 291 A; 475 C; 428 G; 243 T; 0 other;
XX
XX      Alignment Scores:

```

```

Pred. No.: 4.35e-05 1437
Score: 163.00
Percent Similarity: 49.26%
Best Local Similarity: 36.03%
Query Match: 6.21%
DB: 14
Gaps: 6

```

US-10-088-045-2 (1-502) x AAQ48234 (1-1437)

```

Qy      369 LeuProGlyGlyAlaProGlnAlaIleAspThrGlnGlnAlaProIleThrProAlaPro 388
Db      916 GTGCCCGAAGGCAAGAACCTTCGT-----CAGATGACGGTCAGACCGCTCTG 966
Qy      389 AlaAsnLeuGlnGlnInleValAlaGlyValAlaThrTyraSerArgCyAlaAla 408
Db      967 CTGAACAAGGGTGGCCAGGCAACGAGGTGGGAAGTCTATTCGACAACTGGCCATC 1026
Qy      409 CyAlaInglInProAspGlyLysGlyValProAsnAlaPheProProLeuAla---AAsSer 427
Db      1027 TGCCACATGAAGATGATGACGGCGTCAACCGCATGTTCCCGCTGGCAGGCAACCG 1086
Qy      428 AspTyrlLeuAlaAspHlaAlaArgAlaAlaSerlleValAlaAnglLeuSerGly 447
Db      1087 GTCTGATTAACGACGATCCGACGTGCTCGCCAAATGCTGAGCTTCGT-----GGC 1140
Qy      448 LysIleThrValAlaAnglValAnglInTyrgluserVal---MetProAlaIleAla----- 464
Db      1141 ATCTGCCCCCGACCAACGCGACCGCTGCTGTCATGCCCGGGTTCAAGATCAC 1200
Qy      465 LeuSerAspGlnGlnIleAlaAsnValIleThrTyrlLeuAsnSerPheGlyAsnLys 484
Db      1201 CTCTCCGACCAAGAGATGGCGATGTCGTAACCTTTATGGCGAAGGCTGGGCAACAC 1260
Qy      485 Gly---GlyGlnLeuSerAlaAspAspValAlaLysAlaLysThr 499
Db      1261 GCACCGGAACCGTCTCGCTCCGATATCCAGAACTGCGCACAGCG 1308

```

RESULT 23

ABA96893 standard; DNA; 1437 BP.

```

ID      ABA96893 standard; DNA; 1437 BP.
AC      ABA96893;
DT      07-MAY-2002 (first entry)
XX
XX      Gluconobacter oxydans sorbitol dehydrogenase subunit 2 DNA, SEQ ID:7.
XX
XX      Sorbitol dehydrogenase; SDH; subunit 2; Gluconobacter suboxydans;
XX      KM coenzyme PQQ dependent; haem group; ds.
XX
XX      Gluconobacter oxydans.
XX
XX      KR98069057-A.
XX
XX      26-OCT-1998.
XX
XX      26-FEB-1997; 97KR-0005929.
XX
XX      26-FEB-1997; 97KR-0005929.
XX
XX      (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
XX
XX      Chol US, Lee SG, Lee EH;
XX
XX      WPI; 1999-608154/52.
XX
XX      Sorbitol dehydrogenase of Gluconobacter suboxydans and gene thereof -
XX      Claim 11; Page 10; 18pp; Korean.
XX
XX      The invention relates to a sorbitol dehydrogenase (SDH) from
XX      CC Gluconobacter oxydans (referred to as Gluconobacter suboxydans in the
XX      CC specification) and the gene encoding it. The sorbitol dehydrogenase

```


SQ Sequence 1437 BP; 296 A; 442 C; 418 G; 281 T; 0 other;

US-10-088-045-2 (1-502) X AAC83154 (1-1437)

QY 347 -----IleLeuLysValGluGlyGluGluAsnHisGluIleTyrSe 360

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

REPORT 33

AAQ14178
TD 11014170 Standard BYA 2004 PD

XX 2A014179

XX
10 JAN 1963 (Est. not established)

XX
XX
XX

33

XX
XX
Czechoslovak Republic

XX
XX
XX

FT	CDS	1
FT		1

XX
TMO3310100 3

XX
3
3
3
3
1
2
2
2
.

XX 11-TN-1000. 00TB-000E370
DE

XX 11 JAN 1966 : 00TD 0005370

XX (TOWN) TOWNSHIP

XX
DB WBT 1001 314E07/43

DR P-PSDB; AAR14306.

PT DNA encoding polypeptide with ascorbate oxidase activity - derived from *Clavaria* mushroom used to produce recombinant enzyme in high

PT yield.

PS Claim 1, Fig 2, BPP, Japanese.

XX The sequence was obt'd. from a clone isolated from a library prepd.
CC from RNA extracted from the fruit of *C. moschata*. The DNA can be
CC used to produce the enzyme in high yields.

XX Sequence 2064 BP, 574 A, 444 C, 493 G, 553 T, 0 other;

Alignment Scores:

Pred. No.:	Length:	2064
Score:	7.07e-05	163.00
Percent Similarity:	37.69%	Conservative: 44
Best Local Similarity:	24.63%	Mismatches: 130
Query Match:	6.21%	Indels: 80
DB:	12	Gaps: 19

US-10-088-045-2 (1-502) x AAO14178 (1-2064)

```

QY 101 ValGluTyrGlnPheThr-----PheGlyGly 110
DB 118 GTCGAGTACATGTTTGGGCTCCTGATGTGTAATAAATATCGTATGGGAATCAACGGC 177
QY 111 GlnValProGlyGlnMetIleArgValArgGlyAspThrIleGluValGlnPheSer 130
DB 178 CAGTTTCCCTGAGCACTACGATTCAGGCCAACGCTGTGTAATACTGTTGTTGCTGAT 237
QY 131 Asn-----HisProAspSerLysMetProHisAsnValAspPheHisAlaIleThrGly 148
DB 238 AATTAAGCTCCACCTGACAGAGTGTCTGATTCAC-----TGGCATGGAATTTTGCA 288
QY 149 ProGly-----GlyGlyAlaGluAlaSerPheThrAla-----ProGly 161
DB 289 CGAGGAATCCGTCGCTGATGTCATGCTTCATCTCCAGTGTCTATTAACTGCTGCT 348
QY 162 HisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHisCysAla 181
DB 349 GAACGCTTCTTACAACTTTCACCTGTTGTAACCCCTGGAGCGTTTCTATCAT----- 402
QY 182 ValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlnPro 201
DB 403 ---GGTCATTTGGGAATGCAAAAGATGCGAGGGTATATGAGATCTTGTATGATGATCCT 459
QY 202 LysGlnGlyLeuProLys-----ValAspLysGluTyrTyrValMetGlnGlyAsp 218
DB 460 CCACAAGAGAAAAGAGCCATTCATTTATGACGAGAAATACACTATTATCACTAGTAT 519
QY 219 PheTyrThrLysGlyLysTyrGlyGln---GlyLeuGlnProPheAspMetGlyLys 237
DB 520 TGTGTCATCAAAAGTATTCATTAACAGAAAGTTGCTTATGC-----TCCAAA 567
QY 238 AlaIleArg-----GluAspAlaGluTyrValValPheAsnGly----- 250
DB 568 CCTATTTCGTGATCGGTGAGGCTCAACCATATCTTAAATGGAAGAGCGCAGTTGCAT 627
QY 251 ---SerValGlyAlaLeuThrGlyGluAsn-----AlaLeuLysAlaLysValGlyGlu 267
DB 628 TGTTCATATGCACTAAATACGACAGTAATTAGAACCATGTAATAAGAACTGA 687
QY 268 ThrValArgLeuPheValGlyAsnGlyLysProAsnLeuThrSerSerPheHisValIle 287
DB 688 CCGTCGCACTTATATC-----TTCCATGTTATG 717
QY 288 GlyIleIlePheAspLysValHisPheGlnGlyLysGlyGluAsnHisAsnIleGln 307
DB 718 CCCAAAAGACTTACGATTAAGAAAT-----GCAAGT 750
QY 308 ThrThrIleLeuProIleArgGlyAlaIleIleThrGlu-----PheLysValAsp 324
DB 751 ACCACTGCTTTAGCGCCCTCAACTTGGCATTTGGGAATCAACCATTTGTTGGTGA 810
QY 325 ValProGlyAspTyrVal-----LeuValAspHisAlaIlePheArgAlaPhe 340
DB 811 GCCAGCGTAATTAACGTCACCAACCATTTTACACCTCCGACATCGACATTTATTCGGCGAG 870

```

```

QY 341 AsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsnHisGluIleTyr--- 359
DB 871 TCTTACTCGCTCTCATACACGACGCAAAACCATGCGAATACACTAGGCTCCGTC 930
QY 360 -----SerHisLysGlnThrAsp-----AlaValTyrLeuPro 370
DB 931 GGCACCCCGCGAGGAGCTCCCAACACACCGCGGATTTGACCTCTCAACTACCTAACCC 990
QY 371 GlnGlyAlaProGlnAlaIleAspThrGlnGluIleProLysThrProAla 387
DB 991 AATTCGCTCGAAA---TTACCCACTTCTCCGCCCGGAAACCCGCC 1038

RESULT 26
ABA96894
ID ABA96894 strand; DNA; 4830 BP.
AC ABA96894;
AC 07-MAY-2002 (first entry)
DE Gluconobacter oxydans sorbitol dehydrogenase subunit 1-3-encoding DNA.
XX Sorbitol dehydrogenase; SDH, subunit 1; subunit 2; subunit 3;
KW Gluconobacter suboxydans; coenzyme PQQ dependent; haem group; ds.
XX Gluconobacter oxydans.
OS KR98069057-A.
XX 26-OCT-1998.
XX 26-FEB-1997; 97KR-0005929.
XX 26-FEB-1997; 97KR-0005929.
XX (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
XX ChOI US, Lee SG, Lee EH;
XX WPI; 1999-608154/52.
XX Sorbitol dehydrogenase of Gluconobacter suboxydans and gene thereof -
XX Example; Page 14-18; 18pp; Korean.
XX The invention relates to a sorbitol dehydrogenase (SDH) from
XX Gluconobacter oxydans (referred to as Gluconobacter suboxydans in the
XX specification) and the gene encoding it. The sorbitol dehydrogenase
XX comprises 3 subunits of 75 kD, 50 kD and 29 kD, utilizes coenzyme PQQ and
XX has a haem group. The present sequence represents DNA encoding all
XX 3 subunits of the Gluconobacter oxydans sorbitol dehydrogenase of
XX the invention.
XX
SQ Sequence 4830 BP, 1071 A, 1424 C, 1350 G, 985 T, 0 other;

Alignment Scores:
Pred. No.: 0.000221 Length: 4830
Score: 163.00 Matches: 146
Percent Similarity: 34.40% Conservative: 48
Best Local Similarity: 25.89% Mismatches: 187
Query Match: 6.21% Indels: 186
DB: 20 Gaps: 33

US-10-088-045-2 (1-502) x ABA96894 (1-4830)
QY 43 LysThrAlaAsn-----AlaAspAsnAlaIleAspGlnGlnHisGlnGlyLeu 59
DB 2804 AAGCGCGCCAGAGATCTCTACGACAGAGTCAGAGCCGGAAGAAATGAGCAGCGCTC 2863
QY 60 ProValIleAspAlaIleValThrHisAlaProGluValPro-----ProPro 75
DB 2864 CTTAAGCGGCAATTCCTC-----AACGTCCTCAGGGTTCCGTTGCGCT 2908

```

Qy	6	ValAspArgGlnPheHisProAlaLeuValValValMetGluThrValGluValMet	95
Db	2909	GCACGCCGCGGACCATCCGTAAGCGAAACCGTC-----ACGGTGAAGCAATGACG	2955
Qy	96	ArgLeuAlaSerGlyValGluGlyArgPheThrPheGlyGlyGlnAla-----	112
Db	2960	TGACATGCTCAAGGCATTAC-----TCGGAGACAGCTGTATCTGCA	3001
Qy	113	-----ProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer	130
Db	3002	GATGAAACAGGAGTGAATAACCGCGCC-----	3031
Qy	131	AsnHisProAspSerLeuMetProHisAsnValAspPheHisAlaAlaThrGlyProGly	150
Db	3032	-----ACGCGGCTCATGGCAGATGTGTTTCGGTGTCCGCAAGCCGACGAA	3076
Qy	151	GlyGlyValGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLeuVal	170
Db	3077	C-----CTGTATGATGACCC-----CTG	3092
Qy	171	LeuGlnProGlyLeuThrValTyrHis-----CysAlaValAlaPro	184
Db	3093	ATTCAACGGCGGTCCCTAAGCTGGCGCCGCTGTACTGCTGCTGCTGCATACCGCACTA	3152
Qy	185	ValGlyMetHisIleAlaAsnGlyMet-----TyrGly	195
Db	3153	CACGGCCACGCTTTTCTGTGCTGTCTGAGATCAAGACGCGGATCGGACGATCTACTCC	3212
Qy	196	Leu-----IleLeuValGlnProGlyGluGlyLeuProGlyValAspGlySerTyrVal	214
Db	3213	ACCAACATACGCGCTGACCGCAAAATACGGTATCGGCAAC-----TATACA	3257
Qy	215	MetGlnGlyAspPheTyrThrTyrGlyGlySerTyrGlyGlnGlyLeuGlnProPheAsp	234
Db	3258	CTCGAAGAT-----TTCAAGAAAGCG-----	3278
Qy	235	MetGlyValAlaIleArgGluAsp-----	242
Db	3279	ATCCGTAAGGATATCCGCAAGACGCGCGGACGCTTTATCCGCGCATGCCATCTCTGAG	3338
Qy	243	-----LeuGlyValValPheAsnGlySerVal	252
Db	3339	TTCCGCTCGGCTGTCTGATGACGACATCAAGGCCATGTATGCCCTTCTATGCAATGGCGTG	3398
Qy	253	GlyAla-----LeuThrGlyGluAsnAlaLeu	261
Db	3399	AA-GCGCGTCCGCTTCAGAAACAGACGCGGACGATCTCTGCGCGCATGAACGCGCTG	3457
Qy	262	LeuValAlaLeuValGlyGluThrValArg-LeuPheValGlyAsnGlyLeuProAsnLeuThr	281
Db	3458	-----GCCGTGGCCATCTGGCGCGGCGCATGTGTGTCGATCTCAACCCAGGCGCTCGA	3511
Qy	281	rSerSerPheHis-----ValIleGlyGluIlePheAspLeuValHisPheGly	297
Db	3512	CAMAGGCAATCTCCGATCCGGAAGTGGCCGCGGCAATACCTC-----GTGAATGGCCC	3565
Qy	297	uGlyGlyLeuGlyGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaI	317
Db	3566	AGCCCATTTGGCCGAGTGTCAACG-----CCCGTGGCATGGCGCAT	3607
Qy	317	AlleThrGluPheLeuValAspValProGlyAspTyrValLeu-----Val	332
Db	3608	GCAGGTCAAGGGCTATACGGCCAGAGACGGCAAGCTTAACTCTCCGCTGGGCGGACCGCAT	3667
Qy	332	AspHisAlaIle-----PheArgAlaPheAsnGlyValAlaLeuGly-----	346
Db	3668	CGACAACTGATGTCTCCCACTGCGCGTACGACAAATACGACACGGGTGTGGGTGCTGATGTC	3727
Qy	347	-----IleLeuValGlyGluGlyGluAsnHisGluIleTyrSer	360
Db	3728	TGAAGACGACATTGCCAGTTCCTGAAG--ACGGCGCGTATGACCAATTCCTGCGCTTT	3784

QY 360 rHSLvSvGInThrAspAlaVal---Tyr-LeuProGluGlyValProGlnAlaIleAspT 375
 Db 3785 CGGTGCACATGGCTGACGTGTGGCCCTACAGCACCCAGCATGGACATCTGCA 3844
 QY 379 h-rGlnGln1a-----ProLysThrProAlaProAlaLeu---G 392
 Db 3845 CGCAACG-CCCAAGTAACTGAAGACATGCCGCCGTTCCGGAAAGCAAAAACCTGGGTC 3903
 QY 392 InGluGln1aLys-----AlaGlyLysA 400
 Db 3904 AGAGTACCGGCAAGGCCACCGCCCTGCTGAAACCGGTGGCAAGGATGATGACGGCGCAG 3963
 QY 400 lAThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnA 420
 Db 3964 AGGTTTAACTTCACCAACTGTGGCATCTGGCANATAGACATGGCACTGTGTCAACCCGA 4022
 QY 420 lAPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla---Alas 439
 Db 4024 TGTTCCTCCCGCTGTGCTGGCAACCGCGTCATCACCAGACAAATGCAACTCAATGGCCA 4083
 QY 439 eTlleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerV 459
 Db 4084 ACATGTGATGACATTGGC-----GGTATTCTGCTCCGACGAAATACGGCCCATCTGCTG 4137
 QY 459 aI---MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThrT 476
 Db 4138 TTGGCATCGCCGGGCTTCCGCGATCATCTGTCTGACACAGACAGATCGCCGATGTTGAACT 4197
 QY 476 YrThrLeuAsnSerPheGlyIleAsnLysGly---GlyGlnLeuSerLysAlaAspValAlaL 495
 Db 4198 TCATCGCAAGAGCTGGAGGCAACAGCTCCGGAAACCTGTGTGCTCGGATATTCGCA 4257
 QY 495 YsAlaLys 497
 Db 4258 AGCTCCGC 4265
 RESULT 27
 AAC83156
 ID AAC83156 standard; DNA; 4830 BP.
 AAC83156;
 XX
 DT 01-MAR-2001 (first entry)
 XX
 DE DNA encoding Gluconobacter sorbitol dehydrogenase.
 XX
 KW Sorbitol dehydrogenase; SDH; Gluconobacter oxydans; D-sorbitol;
 XX L-sorbose production; 2-keto-L-gulononic acid; ds.
 XX
 OS Gluconobacter oxydans.
 XX
 PN WO200065066-A1.
 XX
 PD 02-NOV-2000.
 XX
 PF 23-APR-1999; 99WO-IB00736.
 XX
 PR 23-APR-1999; 99WO-IB00736.
 XX
 PA (CHOI/) CHOI E.
 PA (RHEE/) RHEE S.
 PA (LEE/) LEE E.
 XX
 PI Choi E, Rhee S, Lee E;
 XX
 DR WPI; 2000-687351/67.
 DR P-PSDB; AAB35987, AAB35988.
 XX
 XX Novel membrane-bound sorbitol dehydrogenase nucleic acid molecule
 PT isolated from Gluconobacter oxydans useful for the fermentative
 PT production of 2-keto-L-gulononic acid and L-sorbose from D-sorbitol -
 XX
 XX Claim 16; Fig 8; 96pp; English.


```

XX AAF26419;
AC
XX
XX 02-MAY-2001 (first entry)
DT
XX
XX Pseudomonas sp heavy metal transporter encoding DNA ORF04246.
DE
XX
XX Heavy metal transporter; iron transporter; transgenic plant;
KW homeostasis regulator; heavy metal ion; trace element; soil pollution;
KW plant growth promoter; plant development; de.
XX
XX Pseudomonas sp.
OS
XX
XX DE19934720-A1.
PN
XX
XX 25-JAN-2001.
PD
XX
XX 23-JUL-1999; 99DE-1034720.
PP
XX
XX 23-JUL-1999; 99DE-1034720.
PR
XX
XX 23-JUL-1999; 99DE-1034720.
PS
XX
XX (TIGR-) TIGR INST GENOMIC RES.
PA (QUIN-) QUINAGEN GMBH.
PA (GESB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
XX WPI; 2001-160596/17.
DR
XX
XX New DNA encoding iron or heavy metal transporters, useful for promoting
PT growth of plants on polluted soil -
PT
XX
XX Claim 3a; Page 14-15; 54pp; German.
PS
XX
XX This invention describes novel DNA sequences (I) that encode iron or
CC heavy metal transporters (II). The invention also describes (1) a
CC recombinant expression vector (III) containing (I); (2) prokaryotic or
CC eukaryotic cells (III) transformed with (I) or (III); (3) production of
CC (II) by culturing (III); (4) (partial) expression products (IV) of (I)
CC and synthetic proteins or peptides with the same sequences; (5)
CC antibodies (ab) specific for (IV); (6) hybridoma cells that produce
CC monoclonal Ab; and (7) transgenic plants that contain (III). The iron or
CC heavy metal transporters encoded by (I) transport heavy metal ions across
CC cell walls and regulate homeostasis of trace elements. (I), and their
CC fragments are useful for: (1) expression of (II); (2) as probes and
CC primers for detection, isolation and amplification of full length cDNA
CC sequences; and (3) producing transgenic plants. (III) are used to promote
CC growth, development and yield of plants, particularly Leguminosae,
CC especially when growing in soil polluted by heavy metals injurious to
CC plants. They also improve homeostasis of iron and trace elements. Host
CC cells that express the iron or heavy metal transporters take up heavy
CC metals, so reduce pollution of soil and release iron and other trace
CC metals which improves soil quality and protect plants against pollutants
CC from the soil.
CC
XX
SQ Sequence 2007 BP; 472 A; 569 C; 594 G; 372 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 8.38e-05 Length: 2007
Score: 162.00 Matches: 82
Percent Similarity: 32.43% Conservative: 49
Best Local Similarity: 20.30% Mismatches: 121
Query Match: 6.17% Indels: 152
DB: 22 Gaps: 17

```

US-10-088-045-2 (1-502) x AAF26419 (1-2007)

```

Qy 110 GlyGlnValProGlyGlnMetIleArgValArgGlnGlyAspThrIleGlnValGlnPhe 129
Db 208 GGCTCCATTCCCGGGCCCATCTGCGTGGCGGCGAAGGCGATACCGTACGCTGCGTGTG 267
Qy 130 SerAn-----HisProAspSerIys-----MetProHisAn 140
Db

```

```

Db 268 CGCAACCGCTTGGACGAGACACCTCCATCCACTGACGACGCGATCATCTCGCGCGGAC 327
Qy 141 ValAsp-----PheHisAlaIleThrGlyProGlyGlyAlaGlu 154
Db 328 ATGACACGGCGCTCCCGGCGCTGACGCTTCATGCG----- 360
Qy 155 AlaserPheThrAlaProGlyIleThrSerThrPheSerPheIleAlaLeuGlnProGly 174
Db 361 -----ATGCGCCGACAGCGGAGATGTATGATGATACAGTTCAAGGTCCGACGAACGCG 411
Qy 175 LeuTyValIleThrIleScyValAlaValAlaProValGlyMethIleAlaLeuGlyMetTy 194
Db 412 ACCCTACTGGTATACAC-----AGCCACTCCGCTTTCAGAGACAGTCCGCGGTGTAT 462
Qy 195 GlyLeuIleLeuValGlnProGlyGlnGlyLeuPro-----LysValAspIleGlyTy 213
Db 463 GGTGCCCTGGTATCATGATGCGAAGAACTGACCGCTTCACTATGACCGGACCTACGTG 522
Qy 214 ValMet----- 215
Db 523 GTCATGTCTACGACGACTGGACGATGAAGACCTGCACGGGTGCTTCCAACTCAAGAG 582
Qy 216 GlyGlyAspPheTyThrIleGlyGly----- 224
Db 583 CAGTCCGACTACTACAACTACCAAGCGCGGTGGCGATTTTCATCAAGATGTGAGC 642
Qy 225 -----TyrlGlyGlnGlyLeuGln 231
Db 643 GAAATGGCTGTGTCGCCGCGGTAGCCGACCGAAGATGTGGCCGAGATGAATAATGAC 702
Qy 232 ProPheAspMetGlyIleValAlaIleArgValAspAlaGlyTyValValPheAsnGlySer 251
Db 703 CCCACGATCTCGCTGATGTGAGCGGCTACACCTACACCTACCTGATG---AACGGCGAG 759
Qy 252 Val-----GlyAlaLeuThrGlyGlyAlaLeuValAlaValGlyGlyIleVal 269
Db 760 GCCCGCGACCGTAACTGACGAGCGC-----GTGTTCAAAGCGCGCGAATAATC 807
Qy 270 ArgLeuPheValGlyGlnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlu 289
Db 808 CGCGTGGCTTCATCAATGCTGCGCCATGACGTATTTGATGTCGAAATTCGCGG--- 864
Qy 290 IlePheAspIleValIlePheGlnGlyGlyGlyGlyAlaLeuValIleGlnIleThrThr 309
Db 865 -----CTGAAGATGACG 876
Qy 310 LeuIleProAlaGlyIle-----AlaAlaIleThrGluPheIleValAsp 324
Db 877 GTTGTGCGCGCGATGGCGACATGTCAACGCGTGTGCGGTCCGACGATTCGCCATCGCC 936
Qy 325 ValProGlyAspTyValIleValAlaPheHisAlaIlePheArgIlePheAsnIleGlyAla 344
Db 937 GTTGCGGAACCTACGACGTC----- 957
Qy 345 LeuGlyIleLeuIleValGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 364
Db 958 -----ATTGTTAGCGCGAGACGACGAGCTTACACCATTTTGTCT----- 999
Qy 365 AspAlaValIleLeuProGlyIleAlaProGlnAlaIleAspThrGlnGlnAlaProGly 384
Db 1000 -----CAATTCATGAT----- 1011
Qy 385 ThrProAlaProAlaAsnLeuGlnGlnGlnIleValAlaGlyIleValAlaThrTyAspSer 404
Db 1012 -----CGCACAGGCTTCTCCAGA 1029
Qy 405 AsnCyAlaAlaCyHisGlnProAspGlyIleGlyValProAlaAlaPhePro---Pro 423
Db 1030 GGCACCTTTAGCGGTGGCGAAGTATGCAAGCTGCGATTCGCGCAATTGATCGGCCCT 1089
Qy 424 LeuAlaAsnSerAspTyIleuAlaAlaAspHisAlaArgAlaAlaSerIleValAlaAsn 443
Db 1090 CTCATCTCATGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1149

```


QY 364 ThrAspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaPro 383
 Db 943020 CCCGACCCGAGTTTCGGCGGATGACATCACTGCGGACGTCGATCCGTGAAATGTTTC 942961
 QY 364 LysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAlaThrThrAsp 403
 Db 942960 ACCGCGGACACGACCTGCAACCTG----- 942937
 QY 404 SerAsnGlyAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAlaPhePro 423
 Db 942936 -----GCCCGCGCCGAACCCACGACGACCTCCG 942907
 QY 424 LeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsn 443
 Db 942906 -----GTCACCTGGGC 942895
 QY 444 GlyLeuSerGlyLysIle-----ThrValAsnGlyAsnGlyTyrGluSerValMetPro 461
 Db 942894 GGCACCATGGCCAGTACGACTGACATCAACGGGGAACCTTACAGCAGCAGCAATCCA 942835

RESULT 30

AA19682/c
 ID AA19682 standard; DNA; 4411529 BP.

AC AA19682;
 DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.

KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 variation; epidemiology; patient treatment; epidemic monitoring; ds.

OS Mycobacterium tuberculosis.

PN US6294328-B1.

PD 25-SEP-2001.

PF 24-JUN-1998; 98US-0103840.

PR 24-JUN-1998; 98US-0103840.

PA (GENO-) INST GENOMIC RES.

PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 WPI; 2001-647261/74.

PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
 determining the nucleotide sequence of the strain at positions in the
 genome corresponding to positions where M. tuberculosis strains CDC
 1551 and H37Rv differ.

XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC H37Rv (AA19682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Alignment Scores:

Pred. No.: 5.28 Length: 4411529
 Score: 158.50 Matches: 104
 Percent Similarity: 31.00% Conservative: 51
 Best Local Similarity: 20.80% Mismatches: 180
 Query Match: 6.04% Indels: 165
 DB: 22 Gaps: 23

US-10-088-045-2 (1-502) x AA19682 (1-4411529)

QY 16 SerAlaLeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProLysSer 35
 Db 944111 AGCGGCTTTCGGCTGACGGCTTTC-----GCCCGAAAGCCACGCGCA 944070
 QY 36 SerThrValAspAlaAlaLysThrAlaAsnAlaAspAsnAla--AlaSerGlnGlu 54
 Db 944069 TCC-----GCCCGCGCGGATGACGCTGCAATGACCGCGCGGCGCGCGCGCG 944016
 QY 55 HisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaProGluValProPro 74
 Db 944015 CACAGTGGGCGCA-----ACGTTACCGCCACGCTGACC-----CCCGAG 943977
 QY 75 ProValAspArgAspHisProAlaLysValAlaValIleMetGlnThrValGluLysVal 94
 Db 943976 CCGCGGAGATCGACCTGGGTGGCCGATCGTACGACCTG----- 943935
 QY 95 MetArgLeuAlaAspGlyValGluTyrGlnPheTyrPheGlyGlyGlnValProGly 114
 Db 943934 -----ACCTACGGCAACACATCCCGCA 943911
 QY 115 GlnMetIleArgValArgGluGlyAspThrIleGluValAlaPheSerAsnHisProAsp 134
 Db 943910 CCACTATCCGGCCACCGTCGGGATGAGATTGCTGCTCGGTGACCAACGCTCGGT 943851
 QY 135 SerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyProGlyGly 152
 Db 943850 GAT-----CCGACGCGGTGATGATGACGCGCATCGCTCGGCAACGATGATGATGCG 943797
 QY 153 AlaGluAlaSerPheThr-----AlaProGlyHisThrSerThrPheSerPheLysAla 170
 Db 943796 ACCGAGCCCGCACCTGCGAACAATCGGCGCGGCTGACTTACGATCGGTTCCCGT 943737
 QY 171 LeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHisIleAla 190
 Db 943736 CCGGATCCGGGACCACTGCGGCGCATCGCGACGTGCTTCGTTCAAGGCGCACAC----- 943683
 QY 191 AsnGlyMetTyrGlyLeuIleuValGluProLysGlyGlyLeuProLysValAspLys 210
 Db 943682 ---GGCTATATCTGCTGCTGCTGCACTGCGACGACGACGACGACGACGACGCGC 943626
 QY 211 GluTyrTyrValMetGlnGlyAspPheTyrThrLysGly----- 223
 Db 943625 GATATGATCATATCTGACGAT--TGAACGAGCGGATCGGGAAGTCCCGCAACAG 943569
 QY 224 LysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla 243
 Db 943568 CTCTACGGGAGAGCTG-----ACGACCGGAACAAACCAACGATGAAACACA 943521
 QY 243 ----- 243
 Db 943520 ACAGGATGCCGGAAGGGAAGCGCTGACAGCACTGTCGCGCGGAGGAGGAGC 943461
 QY 244 -----GluTyrValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla 260
 Db 943460 ATCGCTTACCGCTACTATCTATCAACGGGGAATCCCG-----GTGCGGCGCACGCT 943407
 QY 261 LeuLysAlaLysValGlyGluThrValArgLeuPhePheValGlyAsnGlyGlyProAsnLeu 280
 Db 943406 TTAAAGGCGCAAGCTCGGCGACGATCGGATCGGATCAATCAACAGGCGCGGACACAC 943347
 QY 281 ThrSerSerPheHisValIleGlyGluIlePheAspLysValHisPheGlyGly----- 298
 Db 943346 GCGTTCCGATCGCGCTGGCGGCGATTCGATGACGCTACCCACACGACGAGGTTACCCA 943287

```

QY 299 -----GlyylsglyluasnHisasnIleGln 307
Db 943286 GTGATTCACACGAGTCAGCTCTGATCGCATGGCGAACCCTACGAGTCATG 943227
QY 308 ThrThreuleProAlaIleGlyIle-----AlaAlaIleThrGluPheIleVal 323
Db 943226 GTGACC-----GCCGCTGGCGGCTTCCCTCGTCCGACTCGCGGAGGCGAG-- 943176
QY 324 AspValProGlyAspTyrValIleValAspHisAlaIlePheAspAlaPheAsnIleGly 343
Db 943175 -----AACCGCTGGCGGCGCTGCTGCTCTTACC 943146
QY 344 AlaIleGlyIleLeuIleValIleGlyIleGluIleAsnHisGluIleTyrSerHisIleGln 363
Db 943145 GCGCGCGCG-----AGCCCA 943131
QY 364 ThrAspAlaValIleTyrLeuProGlyIleAlaProGlnAlaIleAspThrGlnIleAlaPro 383
Db 943130 CCGACCCGCGAGTTCCGCGCGAGTAACCTCACTCGAGTGGTACCGGAAATGTTCC 943071
QY 384 LysThrProAlaProAlaAsnLeuGlnIleGlnIleLysAlaIleGlyIleAlaThrTyrAsp 403
Db 943070 ACCGCGCGAAGCACTGCGCAACTG----- 943047
QY 404 SerAsnCySAlaIleAlaCySHisGlnProAspGlyIleGlyIleValProAsnAlaPheProPro 423
Db 943046 -----GGCGCGCGCGAACCACCCAGCACTCCCG 943017
QY 424 LeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaIleAlaIleAlaIleValAlaAsn 443
Db 943016 -----GTCACCTTGGGC 943005
QY 444 GlyLeuSerGlyIleIle-----ThrValAsnGlyAsnGlnTyrGluSerValMetPro 461
Db 943004 GGCACCATGGCCAACTGCACTGGACATCAACGGGAAACCTTACAGACGACCAATCA 942945

RESULT 31
AAK57912
ID AAK57912 standard; DNA; 5187 BP.
AC AAK57912;
XX
DT 15-JUL-1999. (first entry)
XX
DE G. oxydans D-sorbitol dehydrogenase coding sequence.
XX
KW D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulonic acid; precursor;
KM L-ascorbic acid production; ss.
OS Gluconobacter oxydans.
XX
PN W09920763-A1.
XX
PD 29-APR-1999.
XX
PF 13-OCT-1998; 98MO-JP04612.
XX
PR 17-OCT-1997; 97JP-0285280.
XX
PA (FUJII) FUJISAWA PHARM CO LTD.
PI Iehi Y, Noguchi Y, Saito Y, Soeda S, Yoshikawa K;
XX
DR WPI; 1999-302741/25.
XX
PT Gene group for D-sorbitol dehydrogenase, useful for simple
PT large-scale production of L-sorbose or 2-keto-L-gulonic acid as
PT precursor for L-ascorbic acid
XX
PS Claim 21; Page 60-62; 83pp; Japanese.
XX
CC This sequence encodes the D-sorbitol dehydrogenase of the

```

```

CC invention. Cells transformed with a vector containing DNA encoding
CC the dehydrogenase can be used to produce L-sorbose or 2-keto-L-gulonic
CC acid as precursor for simple large-scale L-ascorbic acid production.
XX
SQ Sequence 5187 BP; 1251 A; 1390 C; 1297 G; 1249 T; 0 other;

Alignment Scores:
Pred. No.: 0.00128 Length: 5187
Score: 155.00 Matches: 133
Percent Similarity: 33.17% Conservative: 68
Best Local Similarity: 21.95% Mismatches: 207
Query Match: 5.90% Indels: 199
DB: Gaps: 29

US-10-088-045-2 (1-502) x AAK57912 (1-5187)

QY 15 LeuSerAlaLeuMetLeuSerGlyCySerAsnGlnAlaAspIle----- 29
Db 2667 CTGATGGCTCTGATG--AGTGGTACTGATCTTCAATGACAAAGGGTTACTTCCGAG 2723
QY 30 -----AlaAlaGlnProLysSerSerThrValAspAla 40
Db 2724 TGGCATCCCTCGGCGAGACGATCATGAGAAACAGCCCACTTGGCTTGACGGT 2783
QY 41 AlaAlaLysThrAlaAsnAlaAspAsn-AlaAlaSerGlnGluHisGlnIleGlyIlePro 60
Db 2784 GAGTGGCGGACCCATGACGACGAAACCTGTTGTCSCAGATCAGCGGCTTCTTCCG 2843
QY 60 AlaIleAspAlaIleValThrHisAlaProGluValProProProValAspAlaAspHis 80
Db 2844 GTCGGTACAGCAATA-----TCACCTGACCAATGGCGCGC----- 2880
QY 80 sProAlaIleValIleValIleLysMetGluThrValIleGlyIleValMetArgLeuAlaAspG1 100
Db 2880 ----- 2880
QY 100 ValIleGlyIleGlnPheThrPheGlyIleGlnValPro---GlyIleMetIleArgVal 119
Db 2881 -----TGGCGCTTGGCGTTGCGAGCATCTCTGAAGAAAGAGATGCTTCAT 2924
QY 119 LArgGlyIleAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHis 139
Db 2925 GCGTGAAGGGAATGAACCCGGAATACCGCCCTTTCTCCGACAGCTCC----- 2973
QY 139 sAsnValAspPheHisAlaIleAlaIleThrGlyProGlyIleGlyIleAlaGluAlaSerPheThrAl 159
Db 2974 -----ATAGCTTCGGGTGCTGCTTCCGGCGCAG--TCAGCGAGGCG 3014
QY 159 aProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyIleuTyrVal----- 177
Db 3015 AGAGGATCAGGACCACTATCAGC-----CGAGCGCCCTATCTGGGCTAC 3059
QY 178 -----TyrHisCySAlaValAlaProValGlyMetHisIleAlaAsnIle 193
Db 3060 AGCAGCGCACTGGCTTCCCTGCAATGCAAAACAGGTGGGCTCCCTTCCGGGCGC----- 3114
QY 193 tTyrGlyIleuIleLeuValGluProLysGlyIleuProLysValAspLysGlyIleTyr 213
Db 3115 -----GGCTTGTATGCTGCTCCCAATGGCGGGAGTGT- CGCGTC-----CAACATTA 3163
QY 213 rValMetGlnIleAspPheTyrThrLys-GlyLysTyrGlyIleGlnIleGlyIleuGlnIlePro 233
Db 3164 CACCCGATCCGGA-----TACGGGAATTGGCAAAATACACGAAAGAGAGATTGGCCAAC- 3216
QY 233 heAspMetGlyIleAlaIleArgGluAsp----- 242
Db 3217 --GCTTTCGCAAGGAGTATTCGAGGAGCGAGCTCATCTTACCGGCAATGCTTACA 3274
QY 243 -----AlaGlyIleValIlePheAsnIle 251
Db 3275 CGGCTATTTCGAGATTGGGATTCGAGATTCACGACATTCGATTTGTTACTCATGTCATG 3334
QY 251 erValIleGlyIleuThrGlyIleuAsnAla-----LeuL 262

```

Db 3335 GCGTGGCCCCCTGGCGGACGAGCAATCGGAAGACGACGTAAATTCCTCCCTTCAATATCC 3394
OY 262 YSAIALYValGlyGluThrValAlaGluPheValGly-----AsnGlyIYP 278
Db 3395 GCGCAATGATGATGATCGACTGGAATCTCTGTTCGAGGACCTCCGCCGCCAAGAGGTGATC 3454
OY 278 roAenLeuThrsSerPhe-----HisValIleGlyGluIlePheAspLysValH 295
Db 3455 CTCAGACCTATTCCCAATCGAAGAAAGGCCACTATCTGCAGATCCCTTG----- 3504
OY 295 tAPheGluGlyGlySerGlyGluAsnHisAsnIleGlnThrThrLeuIle----- 311
Db 3505 -----GGACATTGCGGAACCTGTCTATACACCAACGAATTTCTGTATGGCGGACGCA 3556
OY 312 -----P 312
Db 3557 GCACAGTCCCTATTGTGGCGGACCGCGCTCGCTGGTATGTCCCAATCACAAC 3616
OY 312 roAlaGlyGlyAlaAlaIleThrGluPheLysValAspValProGlyAspThrValLeuV 332
Db 3617 CGACGATGAATAGACGGATCGGATCGGCATTGGAGCGAAGACGATGTGTTCAATGACCTGCTA 3676
OY 332 AlaPheHisAlaIlePheAspGlyAlaPheAsnLysGlyValLeuGly----- 346
Db 3677 CAGGCTCCGTCGCGAAGACGTGTCTGAGCGGACGCGCATGATGGCCAGAACCTGTGAACATA 3736
OY 347 --IleLeuLysValGluGlyGluGluAsnHisGlnIleTyrSerHisLysGlnThrAspA 366
Db 3737 GCTTTAGCAAGCTGACAAACGAGATCTCCACGCCATC-----G 3775
OY 366 lAvalTYrLeuProGluGlyAlaProGlnAlaIleAspThrGln----- 380
Db 3776 CCGGCTTAATCCGACAG--ATCCCAAGATCGAGACAGCCAGCAACAGCGCGCTG 3832
OY 381 -----GluAlaProLysThrProAlaP 388
Db 3833 ACCGGTTGGGGTTCCTCCGACGCCCATCGGATCTGCAGAAGCCAAACTTGATGCTG 3892
OY 388 roAlaAsnLeuGlnGluGlnIleLysValGlyLysAlaThrTYrAspSerAsnCYsAlaA 408
Db 3893 AAGATGACCTGTTTCCGATG-----GACGGGAGAGGATGTACTCAACAACTGTGCAG 3946
OY 408 lAcysHisGlnProAspGlyLysGlyValProAsnAlaPheProProLeuAlaAsnSerA 428
Db 3947 CCTGCGATGAACTTGATGTGTGAGAGCGGCCGACCTTCAACGCCCTCTGTCTCTCA 4006
OY 428 sPTYrLeu-----AsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsnG 444
Db 4007 ATGCAAGTAGTGGGTGCACCGGAGGTGACAAATCGATGATGGCC-----ATTGTCAACG 4060
OY 444 lYleuSerGlyLysIleThrValAsnGlyAsnGlnTYrGluSerValMetProAlaIleA 464
Db 4061 GCGTTGATGCG-----ACGACGAATGCT-----CAACAGCTTCTATGCGCGGATTGC 4108
OY 464 lA-----LeuSerAspGlnGlnIleAlaAsnValIleThrTYrT 477
Db 4109 GCCCACTTCGCGATGTACAACGCGCTCAACCGATACGAGATGGCCAAACTCAACCACTATG 4158
OY 477 hTrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspValAlaLysAlaL 497
Db 4169 TCTCCGAGACATTTGGAGATGGCGCATCATGTCAACAGCTCAGAGACGTAAAGGTGCTC 4228
OY 497 ySLyVthrLysPro 501
Db 4229 GTGAAGCGCGGCT 4242

```

RESULT 32
ABT14589
ID ABT14589 standard; DNA; 1125 BP
XX
AC ABT14589;
XX

```

DT 27-FEB-2003 (first entry)

DE Pseudomonas aeruginosa biofilm formation-related gene #31.

XX Gene; ds; biofilm formation modulation; biofilm-associated disease;

KW cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;

KV catheter-associated infection; medical device-associated infection.

XX

OS Pseudomonas aeruginosa.

XX

PV M0200285295-A2.

PD 31-OCT-2002.

XX

PF 19-APR-2002; 2002MO-US12532.

XX

PR 20-APR-2001; 2001US-285190P.

PR 24-OCT-2001; 2001US-344142P.

XX

PA (TOMA) UNIT TOMA RES FOUND.

PA (HARD) HARVARD COLLEGE.

XX

PI Whiteley M, Bangera MG, Lory S, Greenberg EP,

PT WPI: 2003-075501/07.

DR P-PDSB; ABJ18767.

XX

PT Identifying compound capable of modulating biofilm formation by

PT bacteria/bacterial antibiotic resistance, useful for treatment of

PT biofilm associated disease -

XX

PS Claim 1, Page 75-76; 154pp; English.

XX

CC The invention comprises a method for identifying a compound capable of

CC modulating biofilm formation by bacteria. The method of the invention is

CC useful for identifying a compound capable of modulating biofilm formation

CC by bacteria or modulating bacterial antibiotic resistance. The method of

CC the invention is also useful for diagnosing and treating a subject

CC (especially an immunocompromised human) that is afflicted with a biofilm

CC associated disease or disorder, such as: cystic fibrosis; AIDS; middle

CC ear infections; acne; periodontal disease; catheter-associated

CC infections; and medical device-associated infections. The present DNA

CC sequence represents a gene that is used in the invention.

XX

SQ Sequence 1125 BP; 221 A; 362 C; 361 G; 181 T; 0 other;

Alignment Scores:

Pred. No.:	0.000278	Length:	1125
Score:	152.50	Matches:	87
Percent Similarity:	31.8%	Conservative:	40
Best Local Similarity:	21.80%	Mismatches:	110
Query Match:	5.81%	Indels:	162
DB:	25	Gaps:	18

DS-10-088-045-2 (1-502) x ABT14589 (1-1125)

DY 132 HisProASerLysEweCProHisAnValAspHeNiSaLaIaIarHgLYProGlyLY 151

Dd 318 CAACCGAT-----CCACATCTAGACACTTCGAGCCGGAGCTGACGTCAAGGT 365

DY 152 -----GLVALAGLUAlaserPherThrlApRoGlyIsthrSerThrPheSerPhe 166

Dd 369 CACCGGCTACCAAGTAGTGACATTAACAAGTACTCTGG----- 407

DY 169 LysAlaLeuGlNProGlyLeuTYrValYrHisCySaLaVaLaIaProValGIyMechis 188

Dd 408 -----CCAGAACGTGCAGTACTTCAAGCAACTGGCCACC GGAGA GTCA--- 455

DY 189 IleAlaAsnGlyMetTyrcgLYleuILEuValGIuProlYsgLIgLYleuPro-LysVa 208

Dd 453 -----GATCCACA-----CCGCCAGGC 470

DY 208 lAsPlYsgLIUrTYrValMeGInGlyAspPheTYrThrLYsgLYlYsrTYrgLyglU 228

```

Db      471 GAAGGACGACATTACTGCTGAGGTGAC----- 501
Qy      228 nglyleuGlnProPheaspMetGlyAlaIleArgGluaspAlaGluValValPh 248
Db      501 ----- 501
Qy      248 eaanglySerValGlyAlaLeuThrGlyGluaspAlaLeuValValGlyGluTh 268
Db      502 ----- 530
Qy      268 rValArgLeuPheValGlyValanglyGlyProaspLeuThrSerSerPheHisValIleG 288
Db      531 GGTGGCTTCCTGATC-----ACCTCCAGC----- 555
Qy      288 yGluIlePheaspLysValHisPheGlyGlyGlyValGlyGlyValHisValIleGlnTh 308
Db      556 -----GACGTATCCAT-----TC 569
Qy      308 rThrLeuIleProAlaGlyGlyAla-----AlaIleThrGluPhe----- 321
Db      570 CTGGTGGTGCCTGGCTTCGGCGGTCAAGCGGAGCCATCCCGGCTTCGTCAAGAGGC 629
Qy      322 -----LysValAspValProGlyAspTyr----- 329
Db      630 CTGAGACCAAGGTGACGAGCCGCGCATCTATCCGCGCAGTCCGCGAGCTGTGCGCAA 689
Qy      330 -----ValLeuValAspHisAlaIlePheArgAlaPheaspLysGly 343
Db      690 GGAACACGCTTACGCGCATCGGTGTCGAC----- 720
Qy      343 yAlaLeuGlyIleLeuValGlyGlyGlyGluValaspHisGlyIleTyrSerHisLysGly 363
Db      721 -----GTCAAGCCCAAGCGCGAGTTGTGACAGGTGGTGGCCAAAGCCAAAGA 767
Qy      363 nThrAspAlaValTyrLeuProGlyGlyAlaProGlnAlaIleAspThrGlnGluAlaPr 383
Db      768 A--GAGCGCGCGAAGGTCAAGAA-----CTGACCAAGCAAGAGGTGAGAC 809
Qy      383 oLysThrProAlaProAlaaspLeuGlnGlnIleLysAlaGlyLysAlaThrTyrAs 403
Db      810 CAAG-----GAAGAGTTGGTGGCGCGCGCAAGAGTCTTCA 848
Qy      403 pSerAsnCyAlaAlaCyHisGlnProAspGlyLysGlyValProaspAlaPheProPr 423
Db      849 CACCATTTGGCGCCCTGCAACGAGCCGGAAGCCAGGCAATGCTTCCCGGC 908
Qy      423 oLeuAlaAsnSerAspTyrLeuaspAlaAspHisAlaArgAlaAlaSerIleValAlaAs 443
Db      909 GCTGAAGGTTGCAAGATCGTACCGGCGCCAAAGAGACACCACTGGAAGTGTCTTCA 968
Qy      443 nglyLeuSerGly-----LysIleThrValaspGlyAspGlnTyrGlySerValMetProAl 462
Db      969 CGGCGTGGCCGCGGACCGCATGGCGGCTTCGGCAACAG----- 1008
Qy      462 aIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThrLeuaspSerPheG 482
Db      1009 -----CTCAAGAGGTCAAGTCCGCGGTGATACCTTCAAGCGCAACGCTGAGG 1061
Qy      482 yAsnLysGlyGlyLeuSerAlaAspAspValAlaLysAlaLysThrTyrHis 500
Db      1062 CAACGACGATGGCGACATGTGTCAACCCGAAAGACGTGTGCTTCAAGAGCAAA 1116

```

RESULT 33

AAQ20384 standard; DNA; 1404 BP.

```

XX      AC      AAQ20384;
XX      DT      16-APR-1992 (first entry)
XX      DE      ADH complex structural gene (44,000 mol. wt. protein).
XX

```

```

KW      Alcohohl dehydrogenase; acetic acid; fermentation; ss.
XX
OS      Acetobacter alioacetigenes IM-24.
XX
PN      JF03266988-A.
XX
PD      27-NOV-1991.
XX
PF      26-MAR-1990; 90JP-0073440.
XX
PR      26-FEB-1990; 90JP-0042301.
PR      26-MAR-1990; 90JP-0073440.
XX
PA      (NAKA-) NAKANO SUTEN KK.
XX
DR      WPI; 1992-019325/03.
DR      P-PSDB; AAR20193.
XX
PT      Alcohohl dehydrogenase complex structural gene - used in plasmid
PT      and enhancing efficiency of acetic acid fermentation for
PT      transformed acetic acid bacteria
XX
PS      Disclosure; Fig 4(1-3); 21pp; Japanese.
XX
CC      The gene encodes a protein of mol. wt. ca. 44,000. Acetobacter
CC      transformed with the sequence can enhance the efficiency of acetic
CC      acid fermentation. The ADH complex can be easily extracted from the
CC      bacteria and purified and it can be used for the determination of an
CC      alcohol.
CC      See also AAQ20383-84, and -86-88.
XX
SQ      Sequence 1404 BP; 253 A; 448 C; 437 G; 266 T; 0 other;
XX

```

Alignment Scores:

Pred. No.	Length:	1404
Score:	0.000415	
Percent Similarity:	152.00	108
Best Local Similarity:	38.02%	Conservative: 57
Query Match:	5.79%	Mismatches: 165
DB:	13	Indels: 104
		Gaps: 23

```

US-10-088-045-2 (1-502) x AAQ20384 (1-1404)
Qy      113 ProGlyGlnMetIleArgValArg-----GluGlyAspThrIleGluValGln 128
Db      156 CCAATGACAGTCTTACGAGCGGCTTGAATCAAGAGCCGATCGTACGATCTCTC 215
Qy      129 PheSerAsnHis-----ProAsp----- 134
Db      216 CACGAACATCACACCGGACCGGACCTAAGCTATCGGTACACCTTCCGCGAATTGCA 275
Qy      135 ---SerLysMetPro-----HisaspValAspPheHisAlaAla 146
Db      276 CGAAGCGTGGCCATGTATTCGCAAGACGGTTCACCTGTATCCGCGCATGCGGTA 335
Qy      147 Thr-----GlyProGlyGlyValaspGlnIleAspSerPheThrAla 159
Db      336 TCCCTCTTCTTCCGCGATGACGAAGAGACATGCAAGGCGCTGTATCGTATTCATGCA 395
Qy      160 ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyrHis 179
Db      396 TGGGGTGAAGCGCGGTGGCGGACGCGGACAAAGACGCGGACATCTCGGCGCTTGTGCAT 455
Qy      180 CyAlaValAlaProValGlyMetHisIleAlaaspGlyMetTyrGlyLeuIleVal 199
Db      456 GCGCTGGCGCGCTGG-----CATCTGGCCCATGAT-----GTTCTCGCC 494
Qy      200 -GluProLysGlyGlyLeuProLysValAspLysGlyTyrTyrValMetGlnGlyAsp 219
Db      495 TTGCGCGAAGAGCTTACGCGCGCGGCGGACGAGCATCTGGAATGCAAGTGGCATTTA 554
Qy      219 eTyrThrLysGlyLysTyrGlyGlnGlnGlyLeu-----GlnProPheasp 235

```

```

Db 555 TCTGTTACGGGCCCC---GGGCATTGCGGTCGTCTCATACCCCGGTCCTGCCAT 611
Qy 235 T---GluValAlaIleArgGluAspAlaGluTyrValValPheAsnGlySerVal-GlyA 254
Db 612 GCAGGAAAGAGCGCTGCAGCGCTCCGCTGCTGACTTCTGTCCGGTGGCGACCGAT 671
Qy 254 ILeuThrGlyGluAsnAlaLeuValValGlyGluThrValArgLeuPheValG 274
Db 672 GCACAACTGGGTCCGCGGACGTCGCGCA-----ACGATCTGTGCTGGTCT 719
Qy 274 LyAsnGlyGlyProAsnLeuThrSer--SerPheHisValIleGlyGluIlePheAsp 293
Db 720 GGGCGCGCTGGTCCGAGGATGACATCTTACCTTCTGAACTCCGCGCTGACGACCTC 779
Qy 293 ValHisPheGluGlyGlyGlyGluAsnHisAsnIleGlnThrThrLeuIleProAl 313
Db 780 CGCGCGTCTTC---GTTGGCATGGCGCAT----- 804
Qy 313 AgLyGlyAlaAlaIleThrGluPheLyValAspValProGlyAspTyrValLeuValAs 333
Db 805 -----GTGGTGGCATGAGACACCCACTACTTCAACGATGACGACCTG----- 846
Qy 333 PHisAlaIlePheArgAlaPheAsnGlyGlyAlaLeuGlyIleLeuValValGlyGly 353
Db 847 -CACGCCCATCGCGAAG---TACCTGAAGAGCTCCCGCGCTGCGCGCTCACAGGCGCAA 902
Qy 353 uGluAsnHisGluIleTyrSerHisGlyGlnThrAspAlaValTyrLeuProGluGlyAl 373
Db 903 C-----TACACCTTACGATCCGCTCCACCGGACGACCTGCTGGCTGGCT-- 945
Qy 373 AProGlnAlaIleAspThrGlnGluAlaProLySerProAlaProAlaAsnLeuGlnI 393
Db 946 -----AATACCGCGACGCTTCCG----- 963
Qy 393 uGlnIleLeuAlaGlyValAlaThrTyrAspSerAsnGlyAlaAlaCysHisGlnProAs 413
Db 964 -----GGTCTGTATACGTATGTGAAGAAATGCGCCATCTGTCTACCCCTAAGCA 1010
Qy 413 PGIlyLyGlyValProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAs 433
Db 1011 CGGTGGTGGCTGGCCCGCATGTTCCCGCGCTGGCGCAACCGGTTGCTGACCGCA 1070
Qy 433 PHisAlaArgAla---AlaSerIleValAlaAsnGlyLeuSerGlyLySerIleThrValAs 452
Db 1071 GAACCCGACCTCGCTGTGAACGATTTGGCGCATGCT-----GGCGTGGTGGCGCGGAG 1124
Qy 452 nGlyAsnGlnTyrGluSerVal---MetProAlaIle-----AlaLeuSerAspGlnI 469
Db 1125 CAACCTGGGCAACGTCGCCAGTGGCAATGCGCGTTTACAGCAAGATCGCTGCCCGCAGCA 1184
Qy 469 nIleAlaAsnValIleThrTyrThrLeuAsnSerPheGlyAsnLyGly---GlyGlnLe 488
Db 1185 GATTGCTGATGTGTCTCACTTATCCGACACGCTGGGGCAACAGGCGCGCGACCGT 1244
Qy 488 uSerAlaAspAspValAlaLeuValAlaLySerThr 499
Db 1245 TACGGCTGGGATGTTTACCAAGTCGGCGACGACG 1278

```

RESULT 34

AAQ13581 standard; DNA; 1404 BP.

```

ID AAQ13581
XX AAQ13581;
XX
XX 25-MAR-2003 (updated)
DT 09-DEC-1991 (first entry)
XX
XX A.alcoetigenes membrane-bound ADH 44kD sub-unit.
DE
XX alcohol dehydrogenase complex; carboxylic acid production;
KM cytochrome c; ss.
XX
OS Acetobacter alcoetigenes.

```

```

XX EP448969-A.
PN 02-OCT-1991.
PD 26-FEB-1991; 91EP-0102793.
PF 26-MAR-1990; 90JP-0073440.
PR 26-FEB-1990; 90JP-0042391.
XX
PA (NAKA-) NAKANO VINEGAR CO LTD.
PA (NAKA-) NAKANO VINEGAR CO LTD.
PI Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H;
PI Kawamura Y;
DR WPI; 1991-289462/40.
DR P-PSDB; AAR13994.
XX
PT Gene for membrane-bound alcohol dehydrogenase complex - obtd.
PT from Acetobacter alcoetigenes, used for prodn. of enzyme for
PT converting alcohol to acid
XX
PS Claim 4; Page 23-24 and Fig 4; 36pp; English.
XX
CC Total DNA was prepared from A.alcoetigenes MH-24, digested with
CC PstI and SmaI and ligated to pSci-SmaI cleaved pUC18. The ligation
CC mixture was used to transform E.coli DH109. Probes were designed
CC based on the N-terminal amino acid sequence of the 72kD subunit of
CC the ADH complex isolated from A. alcoetigenes (see AAQ13582-Q13584).
CC Fragments encoding the 44kD (cytochrome c) ADH subunit were
CC identified using anti-44kD subunit antibodies in a Western blot.
CC The gene was sequenced. Various experiments suggested that the
CC coding sequence is located immediately downstream of the gene
CC encoding the 72kD sub-unit. See also AAQ13580.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 1404 BP; 254 A; 44% C; 436 G; 266 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 0.00628 Length: 1404
Score: 150.00 Matches: 108
Percent Similarity: 37.79% Conservative: 56
Best Local Similarity: 24.88% Mismatches: 166
Query Match: 5.71% Indels: 104
DB: 12 Gaps: 23

```

US-10-088-045-2 (1-502) x AAQ13581 (1-1404)

```

Qy 113 ProGlyGlnMetIleArgValArg-----GluGlyAspThrIleGluValGln 128
Db 156 CCAATGACAGCTCTTACGACGAGCGGCTGAAATCAAGAGCCGATCGTACGATCTATC 215
Qy 129 PheSerAsnHis-----ProAsp----- 134
Db 216 CACGAACATCAACCGACCGACCGACCTACGATTCGCTGACCTTCCGCAATTCGA 275
Qy 135 ---SerLyMetCPro-----HisAsnValAspPheHisAlaAla 146
Db 276 CGAAGCGGTGGCCATGATTCGCAAGACGGTTCACGCTTATCCGCGCATGCCGTA 335
Qy 147 Thr-----GlyProGlyGlyValAlaGlnAlaSerPheThrAla 159
Db 336 TCCCTCTCTTCCGCGCATGACGAAGAAAGACATGACAGGCGCTGATCCGATTCATGCA 395
Qy 160 ProGlyHisThrSerThrPheSerPheLyAlaLeuGlnProGlyLeuTyrValTyrHis 179
Db 396 TGGGGTGAAGCGCGTGGCGGACCGGACGAAGACGCGGACATCTCGTGGCGCTTGCAT 455
Qy 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuVal 199
Db 456 GCGCTGGCGCGTGG-----CATCTGGCGCATGAT-----GTTCTCGCC 494

```

```

QY 200 -GIuProlYsGIuGIyLeuProlYsValAspLySGLuTYrTYValMecInGIyAspH 219
DB 495 TTCGCGAAGAACTTCAAGCGCGCGCAGCGACCGCGTTCGAAATCCAGATGCGCATTA 554
QY 219 eTYrThrLysGIyLysTYrGIyGIuInGIyLeu-----GlnProPheAspMe 235
DB 555 TCTGCTTACCGGCCCC---GGGCAATTCGGGCTGCTCATACCCCGGCTGCGCATC 611
QY 235 t---GIuLYsAlAlLeaArgLysAspLaGIuTYrValValPheAsnGIySerVal-GIyA 254
DB 612 GCAGGAAAGCGCGTGACCGCTCGGCTGCTTCGATTCCTGTCGCGTGGCGGACCGAT 671
QY 254 lAlLeuThrGIyGIuAsnAlAlLeuLYsAlAlValGIyGIuThrValArgLeuPheValG 274
DB 672 CGCAACTGGGTGCGCGCGCGAGCTGCGCA-----ACGATCTCTGCTGGTGTCT 719
QY 274 lYsAsnGIyGIyProAsnLeuThrSer--SerPheHisValAlLeGIyGIuLLePheAspLy 293
DB 720 GGGCGCGCTGCTCGAGAGATGACATCTACCTTCGAACTCGCGCGCTGTCAGACACATC 779
QY 293 sValHisPheGIuGIyGIyLysGIyGIuAsnHisAsnLLeGIuThrLeuLeuProAl 313
DB 780 CCGCGGTGTC---GCTGCGCATGGCGCAT----- 804
QY 313 aGIyGIyAlAlAlAlLeThrGIuPheLYsValAspValProGIyAspTYrValLeuValAs 333
DB 805 -----GTGGTGCGCATGAGACCGCGCATCTTCCAGCATGACGACCTG----- 846
QY 333 pHisAlAlLePheArgAlAlPheAsnLYsGIyAlAlLeuGIyLYleuLYsValGIuGIyGI 353
DB 847 -CACGCGCATGCGCGAAG--TACCTGAAGACCTGCGCGCGCTGCGCGCGTCAAGGCGAA 902
QY 353 uGIuAsnHisGIuLLeTYrSerHisLYsGIuThrAspAlValTYrLeuProGIuGIyAl 373
DB 903 C-----TACACCTACGATCCGCTCCACCGCGCAACATGCTGCTCGGCT-- 945
QY 373 aProGIuAlAlLeAspThrGIuInGIuAlAProLYsThrProAlAProAlAsnLeuInGI 393
DB 946 -----AATACCGCGCAGCGTTCG----- 963
QY 393 uGIuLLeLYsAlAGLYsAlAlThrTYrAspSerAsnCYsAlAlACyHisGIuInProAs 413
DB 964 -----GGTGCTGATACGATGTGAAGATGCGCATGTGTACCGGTAAACA 1010
QY 413 pGIyLYsGIyValProAsnAlAPheProProlLeuAlAsnSerAspTYrLeuAsnAlAs 433
DB 1011 CGGTGCTGCGTGGCGCGCGCATGTTCCGCGCTGCGTGGCGACCGCGTTCGTGACCGA 1070
QY 433 pHisAlArgAlA--AlAserILEValAlAsnGIyLeuSerGIyLYsILEThrValAs 452
DB 1071 GAACCGGACCTCGCTGCTGTAAGTATGGCGATGGT-----GGGCTGCTGCGCGCGAG 1124
QY 452 nGIyAsnGIuTYrGIuSerVal---MecProAlAlLe-----AlAlLeuSerAspInGI 469
DB 1125 CAACTGGGACCGCTCGCGAGTGCAGATGCGCGGTTCACGACATGCTGTCGCCCGACGA 1184
QY 469 nILEAlAsnValAlLeThrTYrThrLeuAsnSerPheGIyAsnLYsGIy---GIyGlnLe 488
DB 1185 GATTCGTGATGTGTCACTTCATCCGACACACCTGGGGCAACAAGCGCGCGCGACCGT 1244
QY 488 uSerAlAspAspValAlAlLYsAlAlLYsLYsThr 499
DB 1245 TACGCGTGGGATGTTCACCAAGCTGCGCGACAG 1278

```

RESULT 35
 AAQ12068
 ID AAQ12068 standard; DNA; 1668 BP.
 AC AAQ12068;
 XX
 DT 30-AUG-1991 (first entry)
 XX
 DE Ascorbate oxidase from Cucumis sativus.

```

XX XX Ascorbate oxidase; transformant; polypeptide; activity; ss.
KM KM Cucumis sativus.
OS
XX Key Location/Qualifiers
PH CDS 1..1665
FT /tag= a
FT /product= ascorbate oxidase
XX
PN JF03108485-A.
PD 08-MAY-1991.
XX
XX PF 19-SEP-1989; 89UP-0244257.
XX
PR 19-SEP-1989; 89UP-0244257.
XX
XX (TOYM ) TOYOBO KK.
PA (NAGS ) NAGASE SANGYO KPL.
XX
DR WPI; 1991-180924/25.
DR P-PSDB; AAR12339.
XX
XX DNA encoding polypeptide having ascorbate oxidase activity -
PT derived from Cucumis sativus, and transformants producing
PT polypeptide in high yields
PS Disclosure; Fig 2; 9pp; Japanese.
XX
XX The DNA is derived from Cucumis sativus and can be used to obtain
CC large amts. of polypeptide having ascorbate activity.
XX
SQ Sequence 1668 BP; 464 A; 411 C; 377 G; 416 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 0.00164 Length: 1668
Score: 146.50 Matches: 40
Percent Similarity: 45.53% Conservative: 16
Best Local Similarity: 32.52% Mismatches: 42
Query Match: 5.58% Indels: 25
DB: 12 Gaps: 6

```

US-10-088-045-2 (1-502) x AAQ12068 (1-1668)

```

QY 101 ValGIuTYrGIuPheThrPhe-----PheGIyGIy 110
DB 37 GTTGAGTACATGTTTGTGTCGCGAGATTGTGTAACAATTGTATGGGATCAACGCG 96
QY 111 GlnValProGIyGIuMetLLeArgValArgGIuGIyAspThrLLeGIuValGIuPheSer 130
DB 97 GAGTTCCTTGAGCCGACGATTAAGCCAAACGCTGGGCGACATGCTGTGTGGAGCTAACT 156
QY 131 Asn-----HisProAspSerLYsMecProHisAsnValAspPheHisAlAlArhGIy 148
DB 157 AACAACTCCACACGTAAGGTGTTATTAT-----TGGCATGAATCTTAACA 207
QY 149 ProGIy-----GIyGIyAlAGlAlAserPheThrAlA-----ProGIy 161
DB 208 CGAGGTACTCTTGCGGTAGGACAGTCTTCATCTCCAGTGCAGATGTCAATTAAACCGGGT 267
QY 162 HisThrSerThrPheSerPheLYsAlAlLeuGIuInProGIyLeuTYrValTYrHisCYsAlA 181
DB 268 GAGACCTTACCTTACCGGTTTGCGTAGATTAAGCTGGGACATATTATTATCAT----- 321
QY 182 ValAlAProValGIyMecHisAlAlAsnGIyMecTYrGIyLeuLLeuValGIuPro 201
DB 322 ---GGCCATTATGAGATGCAAAAGATCGCTGGGTTGTATGATCTTTATATGATGATCCA 378
QY 202 LysGIuGIy 204
DB 379 CCAAGAGGA 387

```



```

RESULT 36
AAH66018
ID AAH66018 standard; DNA; 1533 BP.
XX
XX
XX AAH66018;
XX
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 1053.
XX
XX Colyiform bacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127668.
XX
XX 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOSYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX P1 Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
XX DR P-PSDB; AAG90799.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PT
XX
XX Claim 8; SEQ ID NO: 1053; 246bp + Sequence listing; English.
XX
XX
XX The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a
XX CC mutant of coryneform bacterium, measuring expression amount and
XX CC analysing the expression profile or expression pattern of a gene derived
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived
XX CC from coryneform bacterium. Coryneform bacteria are useful for producing
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX CC particularly L-lysine. The present sequence is a nucleic acid described
XX CC in the exemplification of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC European Patent Office.
XX
XX
XX Sequence 1533 BP; 321 A; 415 C; 458 G; 339 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. NO.: 0.00372 Length: 1533
XX Score: 142.00 Matches: 99
XX Percent Similarity: 33.72% Conservativity: 47
XX Best Local Similarity: 22.86% Mismatch: 178
XX Query Match: 5.41% Indels: 109
XX DB: Gaps: 20
XX
XX
XX US-10-088-045-2 (1-502) x AAH66018 (1-1533)
XX
XX 29 LyeAlaAlaGlnProlySerSerThrValAAspAlaAlaAlaLysThrAlaAsnAlaAsp 48
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 31 AAAGGGGCGCGGGGCTGCGCAGCAACGGGTGGGCGCCAGGTGCTGAGCGGTTC 90
XX
XX 49 AsnAlaAlaSerGlnGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 65
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 91 TCAGATCATGTGCGTGGTATGGGAGAGCCGCCGACCTTCCTCAATT----- 138

```

OY	66	ValThrHisIslaPProGluValProProProValAsp-----ArgAspHisProAlaLeu	83
Db	139	-----CCACAGCAGATTTAAAGTACCGCGTGAAGGACTTAAAC	174
OY	84	ValValValLysMetGluThrValGluLysValMetArgLeuAlaAspGluValGluTyr	103
Db	175	GTGCACATTTCGCTCGAGAGCTCAGACACTGGGAGAGTACAGATTTCGCGAGATTCAACAG	234
OY	104	GlnIleThrPThrPheGluGluGlnValProGluGlnMetIleArgValArgGluGluAsp	123
Db	235	AAAGACGGGGGATTTCAAATGGACATCTATTGGGGCCGACGTTGGTGTAAGAAAGTGTAT	294
OY	124	ThrIleGluValGlnPheSerAsnHisProAspSer-----	135
Db	295	GACGTCCACGTTGATGATGATTAACAATTGGATGAATATGACCACTGTGCATGGCATGGC	354
OY	136	---LysMetProHisAsnValAspPhe-----HisAlaAlaThrGluProGluGluGly	152
Db	355	ATGAAAGTTGGCCGGAGATGCTGATGATGGTGGTCCGACATCCGATCGGGCGCTGGCCAGACG	414
OY	153	AlaGluAlaSerPheThrAlaProGluYHISerThrPheSerPheLysValAlaGluGln	172
Db	415	TGGTCACCAACGTGACTGTGGCCAAATGATGCACGCACTTTGTGGTATCCACCCGACACT	474
OY	173	ProGluLeuTyrValTyrHisCysAlaValAlaProValGluMetHisIleAlaAsnGly	192
Db	475	CATGGCGCTG-----ACAGGTTTGCATGCTACGCTACGCTGT	507
OY	193	MetTyrGluLeuIleLeuValGluProLysGluGluLeuProLysValAsp-----Lys	210
Db	508	TTGGCGGGGAGATCATTTGTGGAA--GATTAACAACAGACAAGACTGGATCTTCCACGC	564
OY	211	GluTyrTyrVal-----MetGlnGlyAspPheTyrThrLysGly	223
Db	565	GAGTACGGTGGAGCAGATATCCGCTGTTTATATGATCACCGCTTCTTAAGAACGGT	624
OY	224	LysTyrGluGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla	243
Db	625	TCCCTTGATAGAGAAAGCTCCGCG--GATCTTGGCGTGTGGCGGATACCCCACT	678
OY	244	GluTyrValValPheAsnGlySerValGluValLeuThrGluGlnAsnAlaLeuGluAla	263
Db	679	GCCAAATGACATTCACATTCGGGCACTTGGATGCCACACGCCGCGGTTCCGTTCCGCGTG	738
OY	264	LysValGluGluThrValArgLeuPheValGluLysAsnGluGluProAsnLeu-----	280
Db	739	CTCAACGGGCTCCAAATATGACGGTCTCTAT-----AAGTTGGCGCTTTTCA	780
OY	281	---ThrSerSerPheHisValIleGly-----GluIlePheAsp	292
Db	781	GACACGGGCACTTCCCAAGTCATATGGCCAGGCACTTCGCGTTTGTGATGAACCTCAACAC	840
OY	293	LysValHisPheGluGluGluGlyLysGluGluAsnHisAsnIleGlnThrThrLeuIlePro	312
Db	841	CGCACACACTTGGCTATTTGGCGCCAGCGAGCGGAGAAATCGTCGAGACTAGAGCC	900
OY	313	AlaGluGluAlaAlaIleThr-----GluPheLysValAspValPro	326
Db	901	-----GCGAGACGCTGACCTTGGAACTGTGAGCTTTTGAAGCAACTACGCGGCTCCT	954
OY	327	GluAspTyrValLeuValAspHisValaIlePheThrGluPheAsnLysGluValaLeuGly	346
Db	955	GATATATAGTTTCGCGGCCGATTTCCGAGTGTCAATCTTCCAG-----	999
OY	347	IleLeuLysValGluGluGluAsnHisGluIleTyrSerHisLysGlnThrAspAla	366
Db	1000	CTGCTCACTACACCGCGCCTTCCGAT-----GATGCT	1032
OY	367	ValTyrLeuPro-----GluGluAlaProGlnAlaIleAsp	378
Db	1033	GCGCAACACTCTTTGGCGGCGCTGCTGCTGAAATTCACCGAACTGACGTGATGAT	1092

Oy		379 ThrGInGU-u-ALApoLys-----ThirProAlapProAlAsnLeuGlnGlutInleUly 396
Dd	1093 GGCACGTGAAGCGACCTTCATCATGAAACACTTCTCCATCAAGCATT-----	113
Oy	396 sAlAgLyLSAlArThrTyRAsPseRasnCysAlAala 408	
Dd	1139 -----TACAGATGCACATGCAGCGCG 1159	
RESULT 37		
ID	AAF71220 standard; DNA; 1614 BP.	
XX	AAF71220;	
XX	AAF71220;	
DT	30-APR-2001 (first entry)	
DE	Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:165.	
XX		
KW	Corynebacterium glutamicum; homeostasis; adaptation; HA protein;	
KW	fine chemical production; organic acid; proteingenic amino acid;	
KW	nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;	
KW	nucleoside; lipid; saturated fatty acid; unsaturated fatty acid; diol;	
KW	carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;	
KW	diagnosis; Corynebacterium diptheriae; genetic engineering;	
XX	Brevibacterium; environmental condition; ds.	
OS	Corynebacterium glutamicum.	
PN	WO200100842-A2.	
XX		
PD	04-JAN-2001.	
PF	23-JUN-2000; 2000MO-IB00911.	
XX		
PR	25-JUN-1999; 99US-0141031.	
PR	08-JUL-1999; 99DE-1031636.	
PR	09-JUL-1999; 99DE-1032125.	
PR	09-JUL-1999; 99DE-1032126.	
PR	09-JUL-1999; 99DE-1032127.	
PR	09-JUL-1999; 99DE-1032128.	
PR	09-JUL-1999; 99DE-1032129.	
PR	09-JUL-1999; 99DE-1032226.	
PR	14-JUL-1999; 99DE-1032920.	
PR	14-JUL-1999; 99DE-1032922.	
PR	14-JUL-1999; 99DE-1032924.	
PR	14-JUL-1999; 99DE-1032928.	
PR	14-JUL-1999; 99DE-1032930.	
PR	14-JUL-1999; 99DE-1032933.	
PR	14-JUL-1999; 99DE-1032935.	
PR	14-JUL-1999; 99DE-1032973.	
PR	14-JUL-1999; 99DE-1033002.	
PR	14-JUL-1999; 99DE-1033003.	
PR	14-JUL-1999; 99DE-1033005.	
PR	14-JUL-1999; 99DE-1033006.	
PR	31-AUG-1999; 99DE-1041378.	
PR	31-AUG-1999; 99DE-1041379.	
PR	31-AUG-1999; 99DE-1041390.	
PR	31-AUG-1999; 99DE-1041391.	
PR	03-SEP-1999; 99DE-1042088.	
PA	(BADI) BASF AG.	
Pt		
Pt	Pömpeljus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;	
DR	MPJ, 2001-061974/07.	
DR	P-PsDB; AAB79105.	
PT	New isolated Corynebacterium glutamicum nucleic acid for production or	
PT	modulation of production of fine chemicals such as amino acids,	
PT	nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins	
or enzymes -		
Claim 3; Page 370-372; 712pp; English.		

XX		AAE71138 to AAF1357 encode the Corynebacterium glutamicum homeostasis and adaptation (HA) proteins given in AAB79023 to AAB79242. The C. glutamicum HA genes (1) can be used in vectors for expression in host cells and production of fine chemicals, such as, an organic acid, proteogenic or nonproteogenic amino acid (preferred), purine or pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor, polyelectrolyte or enzyme. The amino acids produced can be lysine, glutamine, glutamate, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, or tryptophan. The fine chemical production can be modulated. The presence of (I) or HA proteins encoded by them are used for diagnosing the presence or activity of Corynebacterium diptheriae. (I) can be used to map the C. glutamicum genome or can be used as markers for genetically engineered Corynebacterium or Brevibacterium. The HA proteins encoded by the (I) are used to maintain homeostasis in C. glutamicum or help the microorganism to adapt to different environmental conditions.
SQ	Sequence 1614 BP; 336 A; 421 C; 493 G; 364 T; 0 other;	
XX		
Alignment Scores:		
Pred. No.:	0.00399	Length: 1614
Score:	142.00	Matches: 99
Percent Similarity:	33.72%	Conservative: 47
Best Local Similarity:	22.86%	Mismatches: 178
Query Match:	5.41%	Indels: 109
DB:	22	Gaps: 20
US-10-088-045-2 (1-502) x AAF1220 (1-1614)		
OY	LysAlaIacInProLysSerSerThrValAAspAlaAlaLeuThrAlaAsnAlaAsp	48
Dd	89 AAAGGGGCGGGGTCTGGCAGCAACGGTGGGGTGCGCAGGTCTGGCGGTTCC	148
OY	49 AsnAlaLasereIngluhIsingluYgu-----LeuprovalIIlaAspAlaIle	65
Dd	149 TCAGATGATGGGTGGATTATGGGGGAGGCCGCAGCATTCGCTATT-----	196
OY	66 ValThrlsAlaProgluValProporProValAsp-----ArgAspHisProAlaIys	83
Dd	197 -----CCACCAGCAATTTAGTACSCCGTAGSAGCATTAAC	233
OY	84 ValValValLysMetGluThrValGlusValImetArgLeuAlaAspGIyValGluTyx	103
Dd	233 GTGACATTGGCCCTGGAGGCTCAGACTGGCGGAGATCAGATTTTSCCGAGTCAACAAG	292
OY	104 GlmPheTrpThrPheGluYglYglInValProGIyGlmMetIIeAsgValArgGluGIyAsp	122
Dd	293 AAGACGGGGGTTTTCAATGGACACTAATTTGGGGCGCAGATTCGGTGGAAMAAGGTGAT	352
OY	124 ThrIlseIuValGlnHiseSerAsnHisProAspSer-----	135
Dd	353 GACGTCAACCTTATGATGATAAACAAATTTGGATGAATGACCACTGTGCACCTGGCAGTGC	411
OY	136 ---LysMetProHisAsnValAspPhe-----HisAlaAlaThrGluProGIyGIyGIy	152
Dd	413 ATGAAGTTGGCGCGGCGGATTCGTATGCTGTGCTCCGCACTCACGATCGGGCCCTGGGAGAAG	472
OY	153 AlaGluAlaSerPheThrAlaProGIyHisThrSerThrPheSerPheLysAlaLeuGln	172
Dd	473 TTGTCACCAACGTGGACTGTGGCCCAATGATGCACGACCACTTTGGTGATCAACCCGACACT	533
OY	173 ProGILyLeuThrValYthrHisCyAlaValAlaAlaProValIGlYmeCHisAlaLaangLy	192
Dd	533 CATGGCGCTG-----ACAGGTTTGCATGGCTACCGTGGT	565
OY	193 MetTrcIIyLeuIIeLeuValGluProLysGluGluIyLeuProLysValAsp-----Lys	210
Dd	566 TTGGCGGGGATATCATTTGTGA---GATGAAGCAACAGACAAAGCTGGATCTGCCAAGC	622
OY	211 GlutTyrrVal-----MetGlnGlyAspPheTyrrThrLysGly	223

DB 623 GAGTACCGGTGGACGATATATCCCGTGTATTAATGATCAACCGCTTTAGAAAGCGGT 682
 QY 224 LyeTyrGlyGluGlnGlyLeuGlnProPheSerPheGluValAlaIleArgGluAspAla 243
 DB 683 TCCCTTATATGAGGAGACCTTCCC-----GATCTTGCGCTGTTGGCGATACCCCACT 736
 QY 244 GluTyrValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuGlyAla 263
 DB 737 GCCAATGACATTAACAATGGCCACTTTGATGCCACACACGCGCGGTTCGGTTCGGGTG 796
 QY 264 LysValGlyLeuThrValAlaGluPheValGlyAsnGlyProAsnLeu-----280
 DB 797 CTCACACGGCTTCAATATGCGGTCTAT-----AACTTGCGCTTTTCA 838
 QY 281 ---ThreSerPheHisValIleGly-----GluIlePheAsp 292
 DB 839 GACACGGCGACCTTCCAAATGATTCAGGACGCGATTCGGTTCGATGAACCTCAAGAC 898
 QY 293 LysValHisPheGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 312
 DB 899 CGCAGCAGCTTGGCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 958
 QY 313 AlaGlyGlyAlaAlaIleThr-----GluPheLysValAspValPro 326
 DB 959 -----GGCGAGGAGCTCACCTTGGAACTGTAGATTGTTGAGGACAACTACGCGCTCCCT 1012
 QY 327 GlyAspTyrValLeuValAlaPheHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGly 346
 DB 1013 GATGATGATGCTCGGCCGAGTTTCGCGATGACGATTCCTTCAG-----1057
 QY 347 IleLeuLysValAlaGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 366
 DB 1058 CTGCTGACCATCAACCGCGCTCCGAT-----GATGCT 1090
 QY 367 ValTyrLeuPro-----GluGlyAlaProGlnAlaIleAsp 378
 DB 1091 GCGCAAGCACTGCTTTCGCGCGCTGCTGTAATCCACCGAACTGACGATGATGAT 1150
 QY 379 ThrGlnGlu-AlaProLys-----ThrProAlaProAlaAsnLeuGlnIleLys 396
 DB 1151 GCCACGTGAACGACCTTCATCATGACACCTTCTCCATCAACGATC-----1196
 QY 396 AlaGlyLysAlaThrTyrAspSerAsnCyAlaAla 408
 DB 1197 -----TACAGATGACATGACGCGCG 1217
 RESULT 38
 ID AAH68527 standard; DNA; 349980 BP.
 AC AAH68527;
 DT 26-SEP-2001 (first entry)
 DE C glutamicum coding sequence fragment SEQ ID NO: 7062.
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis; ds.
 XX Corynebacterium glutamicum.
 OS
 PN EPI108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI, 2001-376931/40.
 DR
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PS
 XX
 XX Disclosure; SEQ ID NO: 7062; 246bp + Sequence Listing; English.
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 349980 BP; 79725 A; 90426 C; 98918 G; 80911 T; 0 other;
 Alignment Scores:
 Pred. No: 5.43 Length: 349980
 Score: 142.00 Matches: 99
 Percent Similarity: 33.72% Conservative: 47
 Best Local Similarity: 22.86% Mismatches: 178
 Query Match: 5.41% Indels: 109
 DB: 22 Gaps: 20
 US-10-088-045-2 (1-502) x AAH68527 (1-349980)
 QY 29 LysAlaAlaGlnProLysSerThrValAlaPheAlaAlaLysThrAlaAsnAlaAsp 48
 DB 101362 AAAGGGGCGCGGGCTGCTGCGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 101421
 QY 49 AsnAlaAlaSerGlnLysGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 65
 DB 101422 TCAGATGATGCTGCTGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 101469
 QY 66 ValThrHisAlaProGluValProProProValAsp-----ArgAspHisProAlaLys 83
 DB 101470 -----CCACCAGCAGATTAGTACGCGGTGAGGAGGATCTACG 101505
 QY 84 ValValAlaLysMetGluThrValGlyGlyValMetArgLeuAlaAspGlyValGlyTyr 103
 DB 101506 GTGCACTTGGCTCGAGGCTGACAGTGGGAGAGTCAATTTTGGCGGATGTCACACAG 101565
 QY 104 GlnPheTyrThrPheGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 123
 DB 101566 AAGACGTGGGTTTCAATGACACTCTTTGGGCGGACGCTGGTGATGAAGAAAGTGAAT 101625
 QY 124 ThrIleGluValGlnPheSerHisProAspSer-----135
 DB 101626 GACGTCCAGCTGATGATGAATTAACAATTTGATGAATGAACACCTGTCACCTGGCAGAG 101685
 QY 136 ---LysMetProHisAsnValAspPhe-----HisAlaAlaThrGlyProGlyGlyGly 152
 DB 101686 ATGAAGTTCGCGCGCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101745
 QY 153 AlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGln 172
 DB 101746 TGGTCAACCACTGACCTGTGCGCAATGATGACGACCACTTTGTGTACACACCGGCACT 101805
 QY 173 ProGlyLeuTyrValTyrHisCyAlaValAlaProValGlyMetHisIleAlaAsnGly 192

D	b		101806	CATGGCCCTG-----	ACAGGTTCGATGGCTACCGTGT	1018938
O	y		193	MctYrGlyLeuIleuValGluProLysGluGlyLeuProLysValAsp-----Lys	210.	
D	b		101839	TTGGCGGGATATCATCTGTGGAA---GATGAAGAACAAGAACGCTGGATCTGCCACGC	1018959	
O	y		211	GluTYrTVAl-----MetGlnGlyAspheTYrThrLysGly	223	
D	b		101896	GAGTAGCGGTGGACGATATTCCGCTGTTTAATGATCAACCGCTTCTTAGAAGACGGT	1019555	
O	y		224	LysTYrGlyGluGlnGlyLeuGlnProPheAspMetGlnLysAlaIleArgGluAspAla	243	
D	b		101956	TCCCTTGATGAGGAAGACTCCCC-----GACTTGGCGCTGTGGGCGCATACCACCAC	1020099	
O	y		244	GluTYrValValPheasnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAla	263	
D	b		102010	GCCAATGGCATTCCCAATCCGACTTTGATGCCACCAACCGCGGGTTCGGTCCCGCTG	1020698	
O	y		264	LysValGlyGluThrValArgLeuPheValGlyAsnGlyLysProAsnLeu-----	280	
D	b		102070	CTCAACGCTCCATATATGGGTTCTAT-----AACTTGGCGTTTCA	1021111	
O	y		281	--ThrsSerPheHisValIleGly-----GluIlePheAsp	292	
D	b		102112	GACACGCCACCTTCCAGTCACTGGCACAGCATCCGGTTTGTGATGACCTCAAGAC	1021711	
O	y		293	LysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThrThrLeuIlePro	312	
D	b		102172	CGCACACCTTGCTATTGGCCAGCGGAGCGGTGGAAATCGTCGTGAGCTAGAGCCC	1022311	
O	y		313	AlaGlyGlyAlaAlaIleThr-----GluPheLysValAspValPro	326	
D	b		102232	----GCGAGAGCGTCACTTGGAAATCTGAGTTTTGAGGACAACTACGGCGTCCCT	1022859	
O	y		327	GlyAspTYrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGly	346	
D	b		102286	GATGATGACTTGTGTGCCGATTTTGGCATGTCAAGATTCTTCCAG-----	1023300	
O	y		347	IleLeuLysValGluGlyGluGluAsnHisGluIleTySerHisLysGlnThrAspAla	366	
D	b		102331	CTGCTCACCATCATCCGGCCCTTCCGAT-----GANTGCT	1023639	
O	y		367	ValTYrLeuPro-----GluGlyAlaProGlnAlaIleAsp	378	
D	b		102364	GCGCAGACACCTGCTTTGGCGGCGGTGCTGTAATTCACCGAACCTGACGTCACTCGAT	1024233	
O	y		379	ThrGlnGlu-AlaProLys-----ThrProAlaProAlaAsnLeuGlnGluInIleLys	396	
D	b		102424	GCCACTGAACGCACTTCACTCACTGAACACCTTCTCCATCAACGAT-----	1024698	
O	y		396	sAlaGlyLysAlaThrTYrAspSerAsnCysAlaAla	408	
D	b		102470	-----TACAGATGACATCGACGCG	102490	
RESULT 39						
ID	AAT63317	standard; DNA; 1588 BP.				
XX	AAT63317;					
DT	07-AUG-1997	(first entry)				
DE	Rhizoctonia solani laccase isozyyme 4.(rs14) gene.					
KM	Blue copper oxidase; laccase; enzyme engineering;					
KW	protein engineering; lignin depolymerisation; dye oxidation; ss.					
OS	Rhizoctinia solani.					
XX						
FH	Key	Location/Qualifiers				
FT	CDS	1135..1137				
FT	/transl_except= pos:1135..1137_-aa:Lys					

FT		/note= "ATC codes for Ile"
FT	CDS	1243..1245
FT		/*tag= b
FT		/transl_except= pos:1243..1245; aa:Ile
FT		/note= "AAA codes for Lys"
PX	PN	M09709431-A1.
XX	PD	13-MAR-1997.
XX	PP	03-SEP-1996;
XX	PR	01-SEP-1995;
XX	PA	(NOVO) NOVO NORDISK BIOTECH INC.
XX	PI	Berka RM, Wahlechner JA, Xu F, Berka R;
XX	DR	WP1; 1997-192906/17.
XX	PS	P-PSDB; AAW16301.
PT		New mutant blue copper oxidase enzymes - having different specific activities to wild-type enzymes, used for e.g. liginin (de)polymerisation or oxidation of dyes
CC		Disclosure; Fig 6A-D; 48pp; English.
CC		The rsl4 gene (AAW16317) encoding Rhizoctonia solani laccase isozyyme 4 (AAW16301) can be subjected to site-directed mutagenesis in order to alter e.g. the specific activity or pH-activity profile of the enzyme, or to improve expression yields. The mutation is a deletion, insertion or pref. a substn. of one or more amino acids at a location no greater than 12 Angstroms, pref. no more than 2.5 Angstroms, from the Type I copper site. The mutant enzyme is expressed in transformed host cells for use e.g. in the polymerisation or depolymerisation of lignin, oxidation of dyes, etc.
XX	SQ	Sequence 1588 BP; 362 A; 478 C; 376 G; 372 T; 0 other;
Alignment Scores:		
Pred. No.:	0.00481	Length: 1588
Score:	141.00	Matches: 103
Percent Similarity:	32.28%	Conservative: 60
Best Local Similarity:	20.40%	Mismatches: 216
Query Match:	5.37%	Indels: 126
DB:	18	Gaps: 20
US-10-088-045-2 (1-502) x AAW16317 (1-1588)		
Dy	59	LauProValIleAspAlaIleValThrHisAlaProGluValProProValAspArg 78
Db	22	CTACCTTGTGTCGGTCGCATCAAC----- 48
Dy	79	AspHisProAlaLysValValValLys-----MetGluThrValGluLysValMetArg 96
Db	49	-----CCGCCCTTGTCCTCGCGCAACTATAAGTTGCACATCAAGAAGCATGTC 102
Dy	97	LauAlaAspGlyValGluTrpGlnThrPheThrPheGlyGluValProGlyGlmec 116
Db	103	GCTCCGGAATGGCTTTCAGGCGCTATGCTGTCGTCGAACGGTTTAGTTCGGACGGTG 165
Dy	117	IleArgValArgGluGlyAspThrIleGluValGlnPheSerAnhis-----Pro 133
Db	163	ATACGGCCCAACAAGGCTGACCTTGCGCATTAATGACAGATTCAACGACGACCT 222
Dy	134	AspSerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyPro 149
Db	223	AGTATGCTGTCGCCAACACGATTCATGCGATGGATTGTTCCAAGCTACTACCGCGAC 282
Dy	150	GlyGlyGlyValGluValSerPheThrAlaPro-----GlyHisThrSer 164
Db	283	GAGGATGGCCCCGATTCGTACCGCATGCCCCTATTGGCGAAAAATTGTCTATACATAC 342

QY 165 ThrPheSerPheValAlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaPro 184
 Db 343 GAGATCCCATGGGGCGGCG---CAACAGGAAACATGATGATCAC-----GCCCAT 390
 QY 185 ValGlyMetHis1LeuAlaLeuGlnGlyMetTyrGly---Leu1LeuValGluPro----- 201
 Db 391 CTGGCAGTCATATATGTCATGATGTCGAGCGCTTCTGATCATCTATGATCCAAACGAC 450
 QY 202 -----LysGluGlyLeuProLysValAspLysGluTyrTyrValMetGlnGlyAspPhe 219
 Db 451 CCACACAGTCGGCGCTACGACGTGATGATCCAGCACAGTACATGATCTTGAGAGACTGG 510
 QY 220 Tyr-----ThyLysGlyLysTyrGlyGluGlnGly----- 229
 Db 511 TACCATACTCCGGACCCGTTCTAGAAACAAATGTTCCGATATATACACGGCTCG 570
 QY 230 LeuGlnProPheAspMetGlyValAla1LeuGlnAspAlaGluTyrVal1ValPheAsn 249
 Db 571 CTCTCTCTGTCGGAGCTCGGCTCTTATCAATGCAAGGCGCTATGTC----- 621
 QY 250 GlySerValGlyAlaLeuThrGlyAlaValAlaLeuLysValAlaGlyGluThrVal 269
 Db 622 -----GGCGGTCCCGCAGTCCCGGCTCAGTATCACTAAACGTGGAACGATAT 675
 QY 270 ArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisVal1LeuGlyGlu 289
 Db 676 CGCTTGGCGGTATCAACGCTTCTGATCGGAGCTGTTACCTTTGATC----- 726
 QY 290 IlePheAspLysValHisPheGluGlyGlyLysGlnAsnHisVal1LeuGlnThr 309
 Db 727 -----GACGACATAGTCTG-----ACT 744
 QY 310 Leu1LeuProAlaGlyGly-----AlaAla1LeuGlnPheLysValAsp 324
 Db 745 GTATTTGAGGCGCATGATCTGTCACACCCCTGCTGATGACCTTCCAGATTAC 804
 QY 325 ValPheGlyAspTyrValLeuValAspHisAla1LeuPheAspLysGlyAla 344
 Db 805 GCTGGAACAAGCTACTCTGATCGTTGAGGCCAACCAACCCCGCACTATGATG 864
 QY 345 LeuGly1LeuLysValGluGlyGluGlnAsnHisGlu1LeuTyrSerHisLysGlnThr 364
 Db 865 CGTGCACCAATGACCGTTGCGAGGCGGAACCACTTGAACCCACCAATGTC 924
 QY 365 AspAlaValTyrLeuProGluGlyAlaProGlnAla1LeuAspThrGlnGlu----- 381
 Db 925 TTTCGGTATTTGCACTACGAGGAGGCGCCACACGCCAACCCACGACGAAACAGCAGT 984
 QY 382 -----AlaPro 383
 Db 985 GCTATCGGTACTGCAAGTGTGAAGAGAACGTGATCCGTCATCAACCTGGCGCTCCG 1044
 QY 384 LysThrProAlaProAlaAsnLeuGlnGln1LeuLysAlaGlyLysAlaThrTyrAsp 403
 Db 1045 GCGCGCTCCCTCCCGAGACGTTTCCCTCAATCTTGCAATTTGGCGCCACACAGTTGAT 1104
 QY 404 Ser-----AsnGlyAlaAlaCysHisGlnProAspGlyLysGlyVal 417
 Db 1105 GGGATTCTTAGGTTACATTATATACATCATCTACAGGCTCT----- 1149
 QY 418 ProAsnAlaPheProLysAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla 437
 Db 1150 -----TCGTTGCCACGCTC----- 1164
 QY 438 AlaSer1LeuAlaAsnGlyLysSerGlyLys1LeuThrValAsnGlyAsnGlnTyrGlu 457
 Db 1165 TTGAAGATTTTGGCAAAATGCGACATGACGCCGATTTACGCCCAATGACACACT 1224
 QY 458 SerValMetProAla1LeuAlaLeuSerAspGlnGln1LeuAlaAsnVal1LeuTyrTyr 477
 Db 1225 ATCGTATTCACACAAATAAAGTTATCGAGCTCAATATATACCGGAGGTGACACACCT 1284

QY 478 LeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspValAlaLysAlaLys 497
 Db 1285 ATTCATCTCCACGGCAT-----GTGTTGATATCTCAATATCATCTC 1326
 QY 498 LysThrLysProAsn 502
 Db 1327 GGTGTACCCGCAAC 1341
 RESULT 40
 AA224235
 ID AA224235 standard; DNA; 1588 BP.
 AC AA224235;
 DT 08-FEB-2000 (first entry)
 DE R. solani laccase isozyme 4 DNA.
 KW Laccase; mutant; lignin polymerization; Kraft; liganosulfate; lignin;
 KM depolymerization; phenol; fruit juice; dye; ss.
 OS Rhizoctonia solani.
 XX US5972670-A.
 XX PN 26-OCT-1999.
 XX PD 09-JAN-1998; 98US-0005397.
 XX PF 30-AUG-1996; 96US-0706037.
 XX PR (FENG/) FENG X.
 XX PA (BERK/) BERKA R M.
 XX PA (WAHL/) WAHLLEITHNER J A.
 XX PI Berka RM, Feng X, Wahlleithner JA;
 XX WPI; 2000-021805/02.
 XX DR P-SDB; AAY50728.
 XX PT Mutant Rhizoctonia laccase enzymes useful for the polymerization and
 PT depolymerization of lignin -
 XX PS Example 9; Column 27-32; 43pp; English.
 XX CC This invention describes novel mutant Rhizoctonia solani laccase enzymes
 CC modified to exhibit altered characteristics (e.g. pH activity profile)
 CC relative to the wild-type enzyme. The mutant enzymes have a number of
 CC commercial and industrial applications. For example, they may be used for
 CC the polymerization of lignin (both Kraft and liganosulfates) in solution
 CC to produce high molecular weight lignin. They may also be used for the in
 CC situ depolymerization of lignin in Kraft pulp to produce low molecular
 CC weight lignin. They may also be used for the polymerization of phenolic
 CC or aniline compounds in liquids (for example apple juice may be treated
 CC with the enzyme to accelerate precipitation of the phenolic compounds in
 CC the juice to stabilize it). Additionally, it may be used to decolorize
 CC dyes and dye precursors by oxidation. The mutant laccase enzymes exhibit
 CC different characteristics to those of the wild-type enzyme. In
 CC particular, they may exhibit an altered pH activity profile relative to
 CC the wild-type oxidase or altered specific activities. Additionally, they
 CC may be produced in large quantities. This sequence encodes the
 CC Rhizoctonia solani laccase isozyme 4 which is described in the method of
 CC the invention.
 SO Sequence 1588 BP; 363 A; 481 C; 373 G; 371 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.00481 Length: 1588
 Score: 141.00 Matches: 103
 Percent Similarity: 32.28% Conservative: 60
 Best Local Similarity: 20.40% Mismatches: 216
 Query Match: 5.37% Indels: 126
 DB: 21 Gaps: 20

US-10-068-045-2 (1-502) x AA224235 (1-1588)

59 LeuProValIleAspAlaIleValIleThrHisAlaProGluValProProProValAspArg 78
22 TCACTTTGGTCTGCTGGCTTCACCC----- 48
79 AspHisProAlaLeuValValValLys-----MetGluThrValGluLysValMetArg 96
49 -----CCGCGCTTGGCTGCTGGCTGGCAACTAAGTTGGACATCAAGAACGTCAATGTC 102
97 LeuAlaAspGlyValGluTrpGlnPheTrpThrPheGlyGlyGluValProGlyGlnMet 116
103 GCTCCCGGATGGCTTTCAGCGCTCTATGCTGCTCCGTCACAGCTTATGTTCTGGCAGCTTG 162
117 TLeaGValArgGluGlyAspThrIleGluValGlnPheSerHis-----Pro 133
163 ATCAGCGCGCCAAAGGGTGACACCTTGGCCCATTAATGTGACAGAACTAACTACAGGACCT 222
134 AspSerLysMetProHisAsnValaAspPheHis-----AlaAlaThrGlyPro 149
223 AGTATGCGCTGGCCCAACACAGATCTATTGGCATGATGTTCCAAGTCACTACCGCGAC 282
150 GlyGlyGlyAlaGluAlaSerPheThrAlaPro-----GlyHisThrSer 164
283 GAGGATGGCGCCGCAATTCGTCAAGCAATGCCATATGGCCAAATTTGTCCTATACATAC 342
165 ThrPheSerPheLysAlaLeuGlnProGlyLeuTrpValTyHisGlyValAlaPro 364
343 GAGATCCCAATTCGGCGGC---CAAAACAGAAACATGTGGATAC-----GCCCAT 390
185 ValGlyMetHisIleAlaAsnGlyMetCysTrpGly---LeuIleLeuValGluPro----- 201
391 CTGGGAGTCATATGTGCATGTGATGGATGGACGCGCTTGGTCACTATATGATCCAAACGAC 450
202 -----LysGlnGlyLeuProLysValAspLysGlyTrpValMetGlnLysAspPhe 219
451 CCACAAAGTCCGCGTACAGCGATGATGTCAGACAGACAGTACTCATGCTTGGAGACTGG 510
220 Tyr-----ThrLysGlyLysTrpGlyGlnGlnGly----- 229
511 TACCACTACCCGGCAACCCGCTTAGAAAAGCAAAATGTTCCGACTAATAACACCGCTCTG 570
230 LeuGlnProPheAspMetGluLysAlaIleArgGluAspAlaGluTrpValIlePheAsn 249
571 CTCTCTCCGTTCCGAGCTCGGGCTCTTATCAATGGAAAGGGCGGTATGG----- 621
250 GlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAlaLysValGlyGluThrVal 269
622 -----GGCGGTCCGCGCACTCCCGGTACAGTATCAACGTAACGTCGGAACGATAT 675
270 ArgLysPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlu 289
676 CGCTTGGCGGTAATCAACCGCTTCTGCTATCGGCTGTTTACCTTTTCGACT----- 726
290 IlePheAspLysValHisPheGlnGlyGlyLysGlyGluAsnHisAsnIleGlnThr 309
727 -----GAAGGACATGTGCG-----ACT 744
310 LeuIleProAlaGlyGly-----AlaAlaIleThrGluPheLysValAsp 324
745 GTCATTGAGGCGGATGGATGCTCTGCACACAGCCCTTGCTGTGTACAGCTTCCAGATTTC 804
325 ValProGlyAspTrpValLeuAsnIleAspHisAlaIlePheAsnValaPheAsnLysGlyAla 344
805 GCTGGACAAACGCTACTCTGTCTGTCATGTTGAAGCCAAACCAACCGCGCCGCACTGCGATT 864
345 LeuGlyIleLeuLysValaGluGlyGluAsnHisGlnIleTrpSerHisLysGlnThr 364
865 CGTGACCAACATGACCGTTGCAGAGGCGGGAACCAATGCAAACTTGGACCCCAACATGTG 924
365 AspAlaValaLysLeuProGlnGlyAlaProGlnAlaIleAspThrGlnGlu----- 381

Db	925	TTTCCGCTATTGGCACTACGAGGAGCGCCCAACGCCAACCAGACCGAAGCAAGGAGAGT	984
Qy	382	-----AlaPro	383
Db	985	GCTATCGGTACTGCAGTCGTTGAAGAGACCTCCATCGCTCATCAACCTGGCGCTCGG	1044
Qy	384	LyethrProAlaProAlaAsnLeuGlnGlnGlnIleLysAlaGlyLysAlaThrTyrAsp	403
Db	1045	GGCGGTCGCGCTCCCGCAGACGTTTCCCTCAATCTTGCAATGGGCGCAGACAGTTGAT	1104
Qy	404	Ser-----AsnCysAlaAlaLysHisGlnProAspGlyLysGlyVal	417
Db	1105	GGGATTCCTAGGTTCACTTAATTAATACATCAAGTAGCAGGCTCTT-----	1149
Qy	418	ProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaSerHisAlaArgAla	437
Db	1150	-----TCGTTGCCACGCTC-----	1164
Qy	438	AlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlu	457
Db	1165	TTGAAAGTTTTGGCAACCAATGCGACGACATGACGCCGATTTACGCCCAAAAGACACACT	1224
Qy	458	SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThr	477
Db	1225	ATCGTATTGCCACACAAATAAGTTATGAGCTCAATTACACCGGAGGTCAGACACCCT	1284
Qy	478	LeuAsnSerPheGlyAsnGlySerGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLys	497
Db	1285	ATTCATCTCCACGGCCCAT-----GTGTTGGATATCGTCAATATCACTC	1326
Qy	498	LyethrIlyProAsn	502
Db	1327	GGTGGTACCCCGAAGC	1341

Search completed: September 5, 2003, 14:35:31
Job time : 13762 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:40:47 ; Search time 29 Seconds

(without alignments)
732.416 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502

Sequence: 1 MSKPTLTKTICALSALML.....NKGQLSDADYAKKTKRN 502

Scoring table: OLIGO

Searched: 328717 seqs, 42310858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: Issued Patents AA:*

- 1: /cgn2_6/pdata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/pdata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/pdata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/pdata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/pdata/1/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/pdata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.4	33	US-09-433-428D-40	Sequence 40, Appl
2	7	1.4	104	US-08-817-441-100	Sequence 100, App
3	7	1.4	172	US-09-134-001C-2971	Sequence 2971, Ap
4	7	1.4	177	US-09-252-991A-18782	Sequence 18782, A
5	7	1.4	180	US-09-173-300-28	Sequence 28, Appl
6	7	1.4	180	US-09-252-991A-22780	Sequence 22780, A
7	7	1.4	215	US-08-311-731A-410	Sequence 410, App
8	7	1.4	224	US-09-082-920-2	Sequence 2, Appl1
9	7	1.4	233	US-09-214-631-7	Sequence 7, Appl1
10	7	1.4	234	US-08-299-567-5	Sequence 5, Appl1
11	7	1.4	238	US-08-240-124-2	Sequence 2, Appl1
12	7	1.4	238	US-08-453-943-2	Sequence 2, Appl1
13	7	1.4	238	US-09-057-121-2	Sequence 2, Appl1
14	7	1.4	238	US-09-358-734-2	Sequence 2, Appl1
15	7	1.4	262	US-09-252-991A-22061	Sequence 22061, A
16	7	1.4	300	US-09-252-991A-27792	Sequence 27792, A
17	7	1.4	312	US-08-821-872-2	Sequence 2, Appl1
18	7	1.4	312	US-09-171-517B-11	Sequence 11, Appl1
19	7	1.4	320	US-09-134-001C-4439	Sequence 4439, Ap
20	7	1.4	326	US-09-154-874-9	Sequence 9, Appl1
21	7	1.4	326	US-08-931-668-9	Sequence 9, Appl1
22	7	1.4	327	US-09-154-874-8	Sequence 8, Appl1
23	7	1.4	327	US-08-931-668-8	Sequence 8, Appl1
24	7	1.4	376	US-09-352-991A-21843	Sequence 21843, A
25	7	1.4	438	US-08-886-886-2	Sequence 2, Appl1
26	7	1.4	438	US-09-044-718-2	Sequence 2, Appl1
27	7	1.4	442	US-09-296-284-26	Sequence 26, Appl1

28	7	1.4	450	US-09-044-718-15	Sequence 15, Appl
29	7	1.4	466	US-08-868-433-35	Sequence 35, Appl
30	7	1.4	466	US-08-744-231-35	Sequence 35, Appl
31	7	1.4	478	US-09-296-284-5	Sequence 5, Appl1
32	7	1.4	518	US-08-836-620A-18	Sequence 18, Appl
33	7	1.4	521	US-08-878-563A-3	Sequence 3, Appl1
34	7	1.4	521	US-09-270-117-3	Sequence 3, Appl1
35	7	1.4	534	US-08-878-563A-1	Sequence 1, Appl1
36	7	1.4	539	US-09-270-117-1	Sequence 1, Appl1
37	7	1.4	539	US-09-328-352-5771	Sequence 5771, Ap
38	7	1.4	580	US-09-252-991A-20407	Sequence 20407, A
39	7	1.4	616	US-09-252-991A-19164	Sequence 19164, A
40	7	1.4	727	US-09-252-991A-26100	Sequence 26100, A
41	7	1.4	962	US-09-614-480-9	Sequence 9, Appl1
42	7	1.4	988	US-09-614-480-2	Sequence 2, Appl1
43	7	1.4	1190	US-09-107-532A-7146	Sequence 7146, Ap
44	7	1.2	9	US-09-005-215-13	Sequence 13, Appl
45	6	1.2	9	US-09-344-040C-117	Sequence 117, Appl
46	6	1.2	10	US-08-453-924-16	Sequence 16, Appl
47	6	1.2	11	US-09-133-062D-28	Sequence 28, Appl
48	6	1.2	11	US-09-133-062D-29	Sequence 29, Appl
49	6	1.2	12	US-08-477-674-1	Sequence 1, Appl1
50	6	1.2	12	US-08-473-791-1	Sequence 1, Appl1
51	6	1.2	12	US-08-316-714-1	Sequence 1, Appl1
52	6	1.2	12	US-08-473-673-1	Sequence 1, Appl1
53	6	1.2	13	US-08-305-871A-18	Sequence 18, Appl
54	6	1.2	13	US-08-411-859-15	Sequence 15, Appl
55	6	1.2	13	US-08-788-822A-22	Sequence 22, Appl
56	6	1.2	14	US-09-068-150-1	Sequence 1, Appl1
57	6	1.2	23	US-08-477-674-11	Sequence 11, Appl1
58	6	1.2	23	US-08-473-791-11	Sequence 11, Appl1
59	6	1.2	23	US-08-316-714-11	Sequence 11, Appl1
60	6	1.2	23	US-08-473-673-11	Sequence 11, Appl1
61	6	1.2	23	US-09-425-638A-82	Sequence 82, Appl
62	6	1.2	23	US-09-425-638A-83	Sequence 83, Appl
63	6	1.2	23	US-09-425-638A-84	Sequence 84, Appl
64	6	1.2	23	US-09-543-004-82	Sequence 82, Appl
65	6	1.2	23	US-09-543-004-83	Sequence 83, Appl
66	6	1.2	23	US-09-543-004-84	Sequence 84, Appl
67	6	1.2	31	US-09-071-224-36	Sequence 36, Appl
68	6	1.2	37	US-09-082-278B-1216	Sequence 1216, Ap
69	6	1.2	37	US-09-315-304B-1216	Sequence 1216, Ap
70	6	1.2	37	US-09-834-784-1216	Sequence 1216, Ap
71	6	1.2	41	US-08-453-924-7	Sequence 7, Appl1
72	6	1.2	41	US-08-129-129-5	Sequence 5, Appl1
73	6	1.2	46	US-08-415-751-13	Sequence 13, Appl
74	6	1.2	47	US-08-469-268A-556	Sequence 556, App
75	6	1.2	47	US-08-488-446-556	Sequence 556, App
76	6	1.2	47	US-08-467-344A-556	Sequence 556, App
77	6	1.2	53	US-08-858-207A-444	Sequence 444, App
78	6	1.2	55	US-08-468-580-11	Sequence 11, Appl
79	6	1.2	55	US-08-468-580-12	Sequence 12, Appl
80	6	1.2	55	US-08-468-580-15	Sequence 15, Appl
81	6	1.2	55	US-08-643-212-10	Sequence 10, Appl
82	6	1.2	55	US-08-643-212-12	Sequence 12, Appl
83	6	1.2	55	US-08-643-212-14	Sequence 14, Appl
84	6	1.2	55	US-08-643-212-16	Sequence 16, Appl
85	6	1.2	55	US-08-643-212-18	Sequence 18, Appl
86	6	1.2	55	PCT-US95-03731-11	Sequence 11, Appl
87	6	1.2	55	PCT-US95-03731-12	Sequence 12, Appl
88	6	1.2	55	PCT-US95-03731-15	Sequence 15, Appl
89	6	1.2	55	PCT-US95-03731-16	Sequence 16, Appl
90	6	1.2	72	US-09-328-352-5771	Sequence 5771, Ap
91	6	1.2	74	US-09-252-991A-22657	Sequence 22657, A
92	6	1.2	74	PCT-US92-00282-24	Sequence 24, Appl1
93	6	1.2	75	US-08-604-985A-2	Sequence 2, Appl1
94	6	1.2	77	PCT-US95-02795A-4	Sequence 4, Appl1
95	6	1.2	79	US-08-904-263A-6	Sequence 6, Appl1
96	6	1.2	79	US-09-434-123A-6	Sequence 6, Appl1
97	6	1.2	79	PCT-US95-06266-132	Sequence 132, App
98	6	1.2	80	PCT-US95-06266-114	Sequence 114, App
99	6	1.2	80	PCT-US95-06266-118	Sequence 118, App
100	6	1.2	80	PCT-US95-06266-120	Sequence 120, App

831	6	1.2	1244	3	US-08-938-291A-5	Sequence 5, Appli	904	6	1.2	4655	3	US-08-652-877-84	Sequence 84, Appli
832	6	1.2	1244	4	US-09-589-619-5	Sequence 5, Appli	905	6	1.2	4655	3	US-08-652-877-86	Sequence 86, Appli
833	6	1.2	1247	4	US-09-404-627-2	Sequence 2, Appli	906	6	1.2	4655	3	US-08-652-877-88	Sequence 88, Appli
834	6	1.2	1270	4	US-09-710-092-10	Sequence 9, Appli	907	6	1.2	4655	3	US-08-652-877-90	Sequence 90, Appli
835	6	1.2	1284	4	US-09-343-494-9	Sequence 9, Appli	908	6	1.2	11877	3	US-09-105-537-6	Sequence 6, Appli
836	6	1.2	1386	4	US-09-358-383C-11	Sequence 11, Appli	909	5	1.0	5	1	US-07-609-716-96	Sequence 96, Appli
837	6	1.2	1386	4	US-09-252-991A-28587	Sequence 28587, A	910	5	1.0	5	1	US-08-460-343B-45	Sequence 45, Appli
838	6	1.2	1409	4	US-09-252-991A-22282	Sequence 22282, A	911	5	1.0	5	1	US-08-398-028B-15	Sequence 15, Appli
839	6	1.2	1426	4	US-09-241-606-2	Sequence 14, Appli	912	5	1.0	5	2	US-08-504-265B-45	Sequence 45, Appli
840	6	1.2	1474	4	US-09-710-092-14	Sequence 2, Appli	913	5	1.0	5	3	US-08-475-411A-96	Sequence 96, Appli
841	6	1.2	1477	1	US-08-038-682-4	Sequence 4, Appli	914	5	1.0	5	3	US-08-478-029A-96	Sequence 96, Appli
842	6	1.2	1477	1	US-08-302-882-4	Sequence 4, Appli	915	5	1.0	5	4	US-08-806-029-11	Sequence 11, Appli
843	6	1.2	1477	2	US-08-530-198-4	Sequence 4, Appli	916	5	1.0	5	4	US-09-841-334A-43	Sequence 43, Appli
844	6	1.2	1477	2	US-08-469-880-4	Sequence 4, Appli	917	5	1.0	6	1	US-08-310-429A-10	Sequence 10, Appli
845	6	1.2	1477	2	US-08-728-470-4	Sequence 4, Appli	918	5	1.0	6	2	US-08-690-734A-9	Sequence 9, Appli
846	6	1.2	1477	2	US-08-617-697-4	Sequence 4, Appli	919	5	1.0	6	2	US-08-374-560-9	Sequence 9, Appli
847	6	1.2	1477	3	US-08-719-664-4	Sequence 4, Appli	920	5	1.0	6	2	US-08-374-560-10	Sequence 10, Appli
848	6	1.2	1477	4	US-09-206-942-71	Sequence 71, Appli	921	5	1.0	6	2	US-08-902-622-4	Sequence 4, Appli
849	6	1.2	1504	4	US-09-364-206-2	Sequence 2, Appli	922	5	1.0	6	3	US-08-742-185-9	Sequence 9, Appli
850	6	1.2	1528	1	US-08-463-092B-6	Sequence 6, Appli	923	5	1.0	6	3	US-08-753-247-28	Sequence 28, Appli
851	6	1.2	1528	2	US-08-462-109A-6	Sequence 6, Appli	924	5	1.0	6	4	US-09-187-859-1526	Sequence 1526, Ap
852	6	1.2	1528	2	US-08-460-907B-6	Sequence 6, Appli	925	5	1.0	6	4	US-09-839-542B-1526	Sequence 1526, Ap
853	6	1.2	1528	3	US-08-463-179A-6	Sequence 6, Appli	926	5	1.0	6	4	US-08-877-605-312	Sequence 312, App
854	6	1.2	1528	3	US-08-461-384B-6	Sequence 6, Appli	927	5	1.0	6	4	US-08-877-605-32	Sequence 322, App
855	6	1.2	1530	3	US-09-556-877-178	Sequence 178, App	928	5	1.0	7	2	US-08-666-473-68	Sequence 68, Appli
856	6	1.2	1530	4	US-09-620-412C-178	Sequence 178, App	929	5	1.0	7	3	US-09-190-964-16	Sequence 16, Appli
857	6	1.2	1530	4	US-09-598-419-178	Sequence 178, App	930	5	1.0	7	3	US-09-103-478-5	Sequence 5, Appli
858	6	1.2	1627	4	US-09-252-991A-28863	Sequence 28863, A	931	5	1.0	7	3	US-09-334-601-45	Sequence 45, Appli
859	6	1.2	1627	4	US-09-071-035-258	Sequence 258, App	932	5	1.0	7	4	US-09-193-931C-5	Sequence 5, Appli
860	6	1.2	1638	4	US-09-071-035-262	Sequence 262, App	933	5	1.0	7	4	US-09-187-859-1529	Sequence 1529, Ap
861	6	1.2	1638	4	US-09-071-035-266	Sequence 266, App	934	5	1.0	7	4	US-09-026-221-5	Sequence 5, Appli
862	6	1.2	1638	4	US-08-494-168-2	Sequence 2, Appli	935	5	1.0	7	4	US-09-839-542B-1529	Sequence 1529, Ap
863	6	1.2	1711	3	US-08-369-822C-10	Sequence 10, Appli	936	5	1.0	8	1	US-08-266-514-13	Sequence 13, Appli
864	6	1.2	1711	3	US-08-582-776C-10	Sequence 10, Appli	937	5	1.0	8	1	US-08-335-198-34	Sequence 34, Appli
865	6	1.2	1711	3	US-08-434-831B-10	Sequence 10, Appli	938	5	1.0	8	1	US-08-44C-206B-2	Sequence 2, Appli
866	6	1.2	1780	1	US-08-769-309A-5	Sequence 5, Appli	939	5	1.0	8	2	US-08-807-030-9	Sequence 9, Appli
867	6	1.2	1780	1	US-08-994-570-5	Sequence 5, Appli	940	5	1.0	8	2	US-08-666-473-72	Sequence 72, Appli
868	6	1.2	1792	2	US-08-962-284-4	Sequence 4, Appli	941	5	1.0	8	2	US-08-473-553A-10	Sequence 10, Appli
869	6	1.2	1833	4	US-08-621-944A-4	Sequence 4, Appli	942	5	1.0	8	2	US-08-654-604-13	Sequence 13, Appli
870	6	1.2	1833	4	US-08-945-567D-4	Sequence 4, Appli	943	5	1.0	8	2	US-09-047-144-6	Sequence 6, Appli
871	6	1.2	1992	4	US-08-621-944A-3	Sequence 3, Appli	944	5	1.0	8	3	US-08-963-825-3	Sequence 3, Appli
872	6	1.2	1992	4	US-08-945-567D-3	Sequence 3, Appli	945	5	1.0	8	3	US-08-817-397B-4	Sequence 4, Appli
873	6	1.2	2032	4	US-09-071-035-458	Sequence 458, App	946	5	1.0	8	4	US-09-242-721A-5	Sequence 5, Appli
874	6	1.2	2032	4	US-09-071-035-462	Sequence 462, App	947	5	1.0	8	4	US-09-500-811-3	Sequence 3, Appli
875	6	1.2	2032	4	US-09-071-035-466	Sequence 466, App	948	5	1.0	8	4	US-09-570-573-3	Sequence 3, Appli
876	6	1.2	2048	4	US-09-268-347-48	Sequence 48, Appli	949	5	1.0	8	4	US-09-548-608-3	Sequence 3, Appli
877	6	1.2	2123	3	US-08-968-685A-10	Sequence 10, Appli	950	5	1.0	8	4	US-09-187-859-1532	Sequence 1532, Ap
878	6	1.2	2289	3	US-09-051-019-2	Sequence 2, Appli	951	5	1.0	8	4	US-09-714-146-4	Sequence 4, Appli
879	6	1.2	2404	4	US-09-134-001C-3464	Sequence 3464, Ap	952	5	1.0	8	4	US-09-839-542B-1532	Sequence 1532, Ap
880	6	1.2	2471	3	US-09-112-450-4	Sequence 4, Appli	953	5	1.0	9	1	US-08-032-846-11	Sequence 11, Appli
881	6	1.2	2471	3	US-09-419-291A-4	Sequence 4, Appli	954	5	1.0	9	1	US-08-213-124-26	Sequence 26, Appli
882	6	1.2	2509	4	US-09-252-991A-16642	Sequence 16642, A	955	5	1.0	9	1	US-08-615-181-96	Sequence 96, Appli
883	6	1.2	2710	2	US-08-568-459A-12	Sequence 12, Appli	956	5	1.0	9	2	US-08-340-283-27	Sequence 27, Appli
884	6	1.2	2710	2	US-08-487-826B-12	Sequence 12, Appli	957	5	1.0	9	2	US-09-047-144-1	Sequence 1, Appli
885	6	1.2	2710	2	US-08-487-826B-12	Sequence 12, Appli	958	5	1.0	9	2	US-08-986-234-47	Sequence 47, Appli
886	6	1.2	2713	5	PCR-US96-01735-1	Sequence 1, Appli	959	5	1.0	9	2	US-08-460-890A-65	Sequence 65, Appli
887	6	1.2	2749	4	US-09-385-223A-4	Sequence 4, Appli	960	5	1.0	9	3	US-08-167-641C-65	Sequence 65, Appli
888	6	1.2	2756	1	US-08-375-709-11	Sequence 11, Appli	961	5	1.0	9	3	US-08-159-339A-213	Sequence 213, App
889	6	1.2	2756	1	US-08-752-923-11	Sequence 11, Appli	962	5	1.0	9	3	US-08-159-339A-139	Sequence 139, App
890	6	1.2	2756	3	US-09-090-793-7	Sequence 7, Appli	963	5	1.0	9	3	US-08-159-339A-210	Sequence 210, App
891	6	1.2	2756	3	US-09-231-899-7	Sequence 7, Appli	964	5	1.0	9	3	US-08-159-339A-211	Sequence 211, App
892	6	1.2	3061	2	US-08-487-826B-14	Sequence 14, Appli	965	5	1.0	9	3	US-08-159-339A-212	Sequence 212, App
893	6	1.2	3111	2	US-08-460-309-4	Sequence 4, Appli	966	5	1.0	9	3	US-08-159-339A-821	Sequence 821, App
894	6	1.2	3111	2	US-08-125-077-4	Sequence 4, Appli	967	5	1.0	9	3	US-09-237-176-1	Sequence 1, Appli
895	6	1.2	3788	4	US-09-336-447A-76	Sequence 76, Appli	968	5	1.0	9	3	US-09-347-503-1	Sequence 1, Appli
896	6	1.2	4545	2	US-08-804-227C-14	Sequence 14, Appli	969	5	1.0	9	3	US-09-457-440-1	Sequence 1, Appli
897	6	1.2	4550	2	US-08-804-227C-8	Sequence 8, Appli	970	5	1.0	9	3	US-08-460-971A-65	Sequence 65, Appli
898	6	1.2	4550	2	US-08-804-198-2	Sequence 2, Appli	971	5	1.0	9	3	US-09-005-215-17	Sequence 17, Appli
899	6	1.2	4551	3	US-09-320-878-1	Sequence 1, Appli	972	5	1.0	9	3	US-08-462-040-65	Sequence 65, Appli
900	6	1.2	4551	3	US-09-141-908-2	Sequence 2, Appli	973	5	1.0	9	4	US-08-474-626-11	Sequence 11, Appli
901	6	1.2	4551	4	US-09-657-440-1	Sequence 1, Appli	974	5	1.0	9	4	US-09-187-859-1535	Sequence 1535, Ap
902	6	1.2	4613	3	US-09-105-537-31	Sequence 31, Appli	975	5	1.0	9	4	US-09-839-542B-1535	Sequence 1535, Ap
903	6	1.2	4654	3	US-08-476-515A-84	Sequence 84, Appli	976	5	1.0	9	6	5473052-28	Patent No. 5473052

977 5 1.0 10 1 US-08-003-966-1 Sequence 1, Appl1
978 5 1.0 10 1 US-07-940-861-3 Sequence 3, Appl1
979 5 1.0 10 1 US-08-214-650-16 Sequence 16, Appl1
980 5 1.0 10 1 US-08-459-512-3 Sequence 3, Appl1
981 5 1.0 10 1 US-08-627-497-7 Sequence 17, Appl1
982 5 1.0 10 2 US-08-572-951-34 Sequence 34, Appl1
983 5 1.0 10 2 US-08-533-298-17 Sequence 17, Appl1
984 5 1.0 10 2 US-08-459-657-3 Sequence 3, Appl1
985 5 1.0 10 2 US-08-460-132-3 Sequence 3, Appl1
986 5 1.0 10 3 US-08-159-339A-145 Sequence 145, App
987 5 1.0 10 3 US-08-602-999A-5 Sequence 5, Appl1
988 5 1.0 10 4 US-08-278-865-5 Sequence 5, Appl1
989 5 1.0 10 4 US-09-500-124-5 Sequence 5, Appl1
990 5 1.0 10 4 US-09-106-872A-11 Sequence 11, Appl1
991 5 1.0 10 5 PCT-US92-02050-3 Sequence 3, Appl1
992 5 1.0 11 3 US-08-602-999A-262 Sequence 262, App
993 5 1.0 11 3 US-08-652-877-24 Sequence 24, App
994 5 1.0 11 3 US-08-476-515A-24 Sequence 24, App
995 5 1.0 11 3 US-09-500-124-262 Sequence 262, App
996 5 1.0 11 4 US-09-500-124-262 Sequence 290, App
997 5 1.0 11 6 5187155-26 Patent No. 5187155
998 5 1.0 12 2 US-08-659-567-8 Sequence 8, Appl1
999 5 1.0 12 2 US-08-538-960-4 Sequence 4, Appl1
1000 5 1.0 12 2 US-08-538-960-4

ALIGNMENTS

RESULT 1
US-09-433-428D-40
; Sequence 40, Application US/09433428D
; Patent No. 6149910
; GENERAL INFORMATION:
; APPLICANT: De Leys, Robert J.
; APPLICANT: Zheng, Jian
; TITLE OF INVENTION: PEPTIDES FOR THE DETECTION OF HIV-1 GROUP O
; FILE REFERENCE: CDS-207
; CURRENT APPLICATION NUMBER: US/09/433,428D
; CURRENT FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 40
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-433-428D-40

Query Match 1.4%; Score 7; DB 3; Length 33;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOQIANV 473
DB 1 DOQIANV 7

RESULT 2
US-08-817-441-100
; Sequence 100, Application US/08817441
; Patent No. 6399294
; GENERAL INFORMATION:
; APPLICANT: CHANEAU, PIERRE
; APPLICANT: CLAVEL, FRANCOISE
; APPLICANT: BORMAN, ANDREW
; APPLICANT: QUILLENT, CAROLINE
; APPLICANT: GUETARD, DENISE
; APPLICANT: MONTAGNIER, LUC
; APPLICANT: DONJON DE SAINT-MARTIN, JACQUELINE
; APPLICANT: COHEN, JACQUES
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR
; TITLE OF INVENTION: SUBTYPE) ANTIGENS
; NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,441
; FILING DATE: 11-JUL-1997
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 95/01391
; FILING DATE: 20-OCT-1995
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9412554
; FILING DATE: 20-OCT-1994
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9502526
; FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03260, 6005-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4400
; TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 100:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 104 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-817-441-100

Query Match 1.4%; Score 7; DB 4; Length 104;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOQIANV 473
DB 83 DOQIANV 89

RESULT 3
US-09-134-001C-2971
; Sequence 2971, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2971
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2971

Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 219 FTTKGY 225
|||||
Db 151 FTTKGY 157

RESULT 4

US-09-252-991A-18782
; Sequence 18782, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18782
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18782

Query Match 1.4%; Score 7; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 146 ATGPGGG 152
|||||
Db 171 ATGPGGG 177

RESULT 5

US-09-173-300-28
; Sequence 28, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173,300
; PRIOR FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 28
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-173-300-28

Query Match 1.4%; Score 7; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 KVEGEEN 355
|||||
Db 58 KVEGEEN 64

RESULT 6

US-09-252-991A-22780
; Sequence 22780, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22780
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22780

Query Match 1.4%; Score 7; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKPTLI 7
|||||
Db 119 MSKPTLI 125

RESULT 7

US-08-311-731A-410
; Sequence 410, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 410:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium leprae

US-08-311-731A-410

Query Match 1.4%; Score 7; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 VIDALVT 67
DB 104 VIDALVT 110

RESULT 8
US-09-082-920-2
Sequence 2, Application US/09082920C
Patent No. 6355469
GENERAL INFORMATION:

APPLICANT: Lam, Kelvin T.
TITLE OF INVENTION: Nucleic Acid Encoding M. Tuberculosis ALGU Protein
FILE REFERENCE: 0342/1CS87USAPPEND
CURRENT APPLICATION NUMBER: US/09/082,920C
CURRENT FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 60/035,391
EARLIER FILING DATE: 1997-01-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 224
TYPE: PRF
ORGANISM: Mycobacteria tuberculosis
US-09-082-920-2

Query Match 1.4%; Score 7; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 488 LSADVA 494
DB 83 LSADVA 89

RESULT 9
US-09-214-631-7
Sequence 7, Application US/09214631
Patent No. 6413730
GENERAL INFORMATION:
APPLICANT: Holland, Sacha
APPLICANT: Mbatulu, Geraldine
APPLICANT: Pawsen, Tony
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
FILE REFERENCE: 11757.23USMO
CURRENT APPLICATION NUMBER: US/09/214,631
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: PCT/CA97/00473
EARLIER FILING DATE: 1997-07-04
EARLIER APPLICATION NUMBER: 60/021,272
EARLIER FILING DATE: 1996-07-05
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 233
TYPE: PRF
ORGANISM: Homo sapiens
US-09-214-631-7

Query Match 1.4%; Score 7; DB 4; Length 233;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
DB 74 GPGGAE 80

RESULT 10
US-08-299-567-5
Sequence 5, Application US/08299567
Patent No. 5747033
GENERAL INFORMATION:

APPLICANT: Davis, et al.
TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL
TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-6707
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,567
FILING DATE: 01-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kempler, Gail M.
REGISTRATION NUMBER: 32,143
REFERENCE/DOCKET NUMBER: REG 290
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-299-567-5

Query Match 1.4%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
DB 74 GPGGAE 80

RESULT 11
US-08-240-124-2
Sequence 2, Application US/08240124
Patent No. 5516658
GENERAL INFORMATION:
APPLICANT: BECKMANN, M. P.
APPLICANT: CERRETTI, DOUGLAS P.
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
TITLE OF INVENTION: RECEPTOR HEK
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1

SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/240,124
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,132
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/114,426
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/109,745
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: SEESE, KATHRYN A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2814-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-240-124-2

Query Match 1.4%; Score 7; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
|||||
Db 78 GPGGAE 84

RESULT 12
US-08-453-943-2
Sequence 2, Application US/08453943
Patent No. 5738844
GENERAL INFORMATION:
APPLICANT: BECKMANN, M. P.
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
TITLE OF INVENTION: RECEPTOR HEK
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,943
FILING DATE: 30-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/240,124
FILING DATE: 09-MAY-1994
APPLICATION NUMBER: US 08/161,132
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/114,426
FILING DATE: 30-AUG-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/109,745
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: SEESE, KATHRYN A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2814-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-943-2

Query Match 1.4%; Score 7; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
|||||
Db 78 GPGGAE 84

RESULT 13
US-09-057-121-2
Sequence 2, Application US/09057121
Patent No. 5969110
GENERAL INFORMATION:
APPLICANT: BECKMANN, M. P.
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
TITLE OF INVENTION: RECEPTOR HEK
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/057,121
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/240,124
FILING DATE:
APPLICATION NUMBER: US 08/161,132
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/114,426
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/109,745
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: SEESE, KATHRYN A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2814-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-057-121-2

Query Match 1.4%; Score 7; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
DB 78 GPGGAE 84

RESULT 14

US-09-358-734-2
Sequence 2, Application US/09358734
Patent No. 6274117

GENERAL INFORMATION:

APPLICANT: BECKMANN, M. P.

TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE

TITLE OF INVENTION: RECEPTOR HER

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

STREET: 51 UNIVERSITY STREET

CITY: SEATTLE

STATE: WASHINGTON

COUNTRY: USA

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: Apple System 7.1

SOFTWARE: Microsoft Word for Apple, Version 5.1a

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/358,734

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/240,124

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/114,426

FILING DATE: 30-AUG-1993

APPLICATION DATA:

APPLICATION NUMBER: US 08/109,745

FILING DATE: 20-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: SESE, KATHRYN A.

REGISTRATION NUMBER: 32,172

REFERENCE/DOCKET NUMBER: 2814-C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 587-0430

TELEFAX: (206) 233-0644

TELEX: 756822

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-358-734-2

DB 78 GPGGAE 84

RESULT 15

US-09-252-991A-22061
Sequence 22061, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 22061

LENGTH: 262

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22061

Query Match

Best Local Similarity 1.4%; Score 7; DB 4; Length 262;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LMSGCS 24
DB 213 LMSGCS 219

RESULT 16

US-09-252-991A-27792
Sequence 27792, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 27792

LENGTH: 300

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27792

Query Match

Best Local Similarity 1.4%; Score 7; DB 4; Length 300;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 435 ARAASIV 441
DB 195 ARAASIV 201

RESULT 17

US-08-821-872-2
Sequence 2, Application US/08821872A
Patent No. 6204004

GENERAL INFORMATION:

APPLICANT: Kaper, James B

APPLICANT: Jarvis, Karen

QY 148 GPGGAE 154

```

; TITLE OF INVENTION: Immunodiagnostic Test for Enterohemorrhagic Escherichia
; FILE OF INVENTION: coli Infection
; FILE REFERENCE: Seq. ID No. 62040044 Ref: 14190K
; CURRENT APPLICATION NUMBER: US/08/821,872A
; CURRENT FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-821-872-2

Query Match
Best Local Similarity 1.4%; Score 7; DB 3; Length 312;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 ADDVAKA 496
    |||||
Db 187 ADDVAKA 193

RESULT 18
US-09-171-517B-11
; Sequence 11, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
; CURRENT APPLICATION NUMBER: US/09/171,517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-171-517B-11

Query Match
Best Local Similarity 1.4%; Score 7; DB 4; Length 312;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 ADDVAKA 496
    |||||
Db 187 ADDVAKA 193

RESULT 19
US-09-134-001C-4439
; Sequence 4439, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4439
; LENGTH: 320
```

```

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4439

Query Match
Best Local Similarity 1.4%; Score 7; DB 4; Length 320;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LMLSGCS 24
    |||||
Db 17 LMLSGCS 23

RESULT 20
US-09-154-874-9
; Sequence 9, Application US/09154874
; Patent No. 6054636
; GENERAL INFORMATION:
; APPLICANT: FADER, GARY MICHAEL
; TITLE OF INVENTION: ISOFLAVONE BIOSYNTHETIC ENZYMES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,874
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/931,668
; FILING DATE: SEPTEMBER 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARIAN, WILLIAM R.
; REGISTRATION NUMBER: 41,173
; REFERENCE/DOCKET NUMBER: BB-1098-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4926
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-154-874-9

Query Match
Best Local Similarity 1.4%; Score 7; DB 3; Length 326;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GALGILK 349
    |||||
Db 107 GALGILK 113

RESULT 21
US-08-931-668-9
; Sequence 9, Application US/08931668
; Patent No. 6521433
; GENERAL INFORMATION:
; APPLICANT: FADER, GARY M.
; TITLE OF INVENTION: CDNA SEQUENCES FROM SOYBEAN THAT ENCODE
; ACTIVITIES ASSOCIATED WITH ISOFLAVONE
```

TITLE OF INVENTION: BIOSYNTHESIS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,668
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MAJARIAN, WILLIAM R.
REGISTRATION NUMBER: 41,173
REFERENCE/DOCKET NUMBER: BB-1098
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4926
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 326 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-931-668-9

Query Match 1.4%; Score 7; DB 4; Length 326;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GALGILK 349
DB 107 GALGILK 113

RESULT 22
US-09-154-874-8
Sequence 8, Application US/09154874
Patent No. 6034636
GENERAL INFORMATION:
APPLICANT: FADER, GARY MICHAEL
TITLE OF INVENTION: ISOPHYLONE BIOSYNTHETIC ENZYMES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,874
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/931,668
FILING DATE: SEPTEMBER 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: MAJARIAN, WILLIAM R.
REGISTRATION NUMBER: 41,173

REFERENCE/DOCKET NUMBER: BB-1098-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4926
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-154-874-8

Query Match 1.4%; Score 7; DB 3; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GALGILK 349
DB 107 GALGILK 113

RESULT 23
US-08-931-668-8
Sequence 8, Application US/08931668
Patent No. 6521433
GENERAL INFORMATION:
APPLICANT: FADER, GARY M.
TITLE OF INVENTION: CDNA SEQUENCES FROM SOYBEAN THAT ENCODE
TITLE OF INVENTION: ACTIVITIES ASSOCIATED WITH ISOPHYLONE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,668
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MAJARIAN, WILLIAM R.
REGISTRATION NUMBER: 41,173
REFERENCE/DOCKET NUMBER: BB-1098
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4926
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-931-668-8

Query Match 1.4%; Score 7; DB 4; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GALGILK 349
DB 107 GALGILK 113

RESULT 24
US-09-252-991A-21843

; Sequence 21843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21843
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21843

Query Match 1.4%; Score 7; DB 4; Length 376;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 264 KVGETVR 270
Db 322 KVGETVR 328

RESULT 25
US-08-886-886-2
; Sequence 2, Application US/0886886
; Patent No. 6107068

; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Delcardayre, Stephen B.
; APPLICANT: Davies, Julian E.
; TITLE OF INVENTION: COENZYME A DISULFIDE REDUCTASE,
; TITLE OF INVENTION: AND INHIBITORS THEREOF USEFUL AS ANTIMICROBIAL
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/886,886
; FILING DATE: 02-JUL-1997

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 6016.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; TELEFAX: 847-938-2623
; TELEX:

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-886-886-2

Query Match 1.4%; Score 7; DB 3; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 436 RAASIVA 442
Db 300 RAASIVA 306

RESULT 26
US-09-044-718-2
; Sequence 2, Application US/09044718
; Patent No. 6391605
; GENERAL INFORMATION:
; APPLICANT: KOSTREMA, Dirk
; APPLICANT: PASAMONTES, Luis
; APPLICANT: TOMSCHY, Andrea
; APPLICANT: van LOON, Adolphus
; APPLICANT: VOGEL, Kurt

; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: MODIFIED PHYTASES
; FILE REFERENCE: Modified Phytases
; CURRENT APPLICATION NUMBER: US/09/044,718
; CURRENT FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: EP 97810175.6
; PRIOR FILING DATE: 1997-03-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Aspergillus terreus
US-09-044-718-2

Query Match 1.4%; Score 7; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 487 QLSADDV 493
Db 223 QLSADDV 229

RESULT 27
US-09-296-284-26
; Sequence 26, Application US/09296284A
; Patent No. 6204040

; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes
; TITLE OF INVENTION: and Methods of Use Thereof
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Glucobacter suboxydans
US-09-296-284-26

Query Match 1.4%; Score 7; DB 3; Length 442;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 LSDOOIA 471

Db 365 LSDQOIA 371

RESULT 28
US-09-044-718-15
Sequence 15, Application US/09044718
Patent No. 6391605
GENERAL INFORMATION:
APPLICANT: KOSTREMA, Dirk
APPLICANT: PASAMONTES, Luis
APPLICANT: TOMSCHY, Andrea
APPLICANT: van LOON, Adolphus
APPLICANT: VOGEL, Kurt
APPLICANT: WYSS, Markus
TITLE OF INVENTION: MODIFIED PHYTASES
FILE REFERENCE: Modified Phytases
CURRENT APPLICATION NUMBER: US/09/044,718
CURRENT FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: EP 97810175.6
PRIOR FILING DATE: 1997-03-25
NUMBER OF SEQ ID NOS: 82
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 450
TYPE: PRT
ORGANISM: Aspergillus terreus
US-09-044-718-15

Query Match 1.4%; Score 7; DB 4; Length 450;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 QLSADVV 493
Db 235 QLSADVV 241

RESULT 29
US-08-868-435-35
Sequence 35, Application US/08868435
Patent No. 6291221
GENERAL INFORMATION:
APPLICANT: Van Loon, Adolphus
APPLICANT: Mitchell, David
TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,435
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/744,231
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kaas, Alan P
REGISTRATION NUMBER: 32142
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-4205
TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 466 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc_feature
LOCATION: 27
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 120
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 207
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 230
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 352
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 376
OTHER INFORMATION: /note="potential N-glycosylation site"
US-08-868-435-35

Query Match 1.4%; Score 7; DB 3; Length 466;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 QLSADVV 493
Db 251 QLSADVV 257

RESULT 30
US-08-744-231-35
Sequence 35, Application US/08744231
Patent No. 6358722
GENERAL INFORMATION:
APPLICANT: Van Loon, Adolphus
APPLICANT: Mitchell, David
TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/744,231
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/424,757
FILING DATE: 18-Apr-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kaas, Alan P
REGISTRATION NUMBER: 32142
REFERENCE/DOCKET NUMBER: Case Docket 9339

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-4205
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 466 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc_feature
LOCATION: 27
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 120
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 207
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 230
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 352
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 376
OTHER INFORMATION: /note="potential N-glycosylation site"
OTHER INFORMATION: /note="potential N-glycosylation site"
US-08-744-231-35

Query Match
Best Local Similarity 1.4%; Score 7; DB 4; Length 466;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 QLSADDV 493
DB 251 QLSADDV 257

RESULT 31
US-09-296-284-5
Sequence 5, Application US/09296284A
GENERAL INFORMATION:
APPLICANT: Choi, Eun-Sung
APPLICANT: Rhee, Sang-Ki
APPLICANT: Lee, Eun-Hae
TITLE OF INVENTION: Glucosyltransferase Substrates Sorbitol Dehydrogenase, Genes
FILE REFERENCE: 1533.0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
NUMBER OF SEQ ID NOS: 87
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 5
LENGTH: 478
TYPE: PRT
ORGANISM: Glucosyltransferase suboxydants
US-09-296-284-5
```

```
Query Match
Best Local Similarity 1.4%; Score 7; DB 3; Length 478;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 LSDOQIA 471
DB 401 LSDOQIA 407
```

```
RESULT 32
US-08-836-620A-18
Sequence 18, Application US/08836620A
Patent No. 5958710
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Orphan receptor
NUMBER OF SEQUENCES: 19
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,620A
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/03933
FILING DATE:
APPLICATION NUMBER: GB 9518272.1
FILING DATE: 08-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9605550.4
FILING DATE: 15-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9607532.0
FILING DATE: 11-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9609576.5
FILING DATE: 08-MAY-1996
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 amino acids
TYPE: amino acid
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-836-620A-18
```

```
Query Match
Best Local Similarity 1.4%; Score 7; DB 2; Length 518;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 148 GPGGAE 154
DB 466 GPGGAE 472
```

```
RESULT 33
US-08-878-563A-3
Sequence 3, Application US/08878563A
Patent No. 5851674
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,563A
```

FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0323 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1203820
US-08-878-563A-3

Query*Match 1.4%; Score 7; DB 2; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 KAAQPKS 35
|||||
Db 318 KAAQPKS 324

RESULT 34
US-09-270-117-3
Sequence 3, Application US/09270117
Patent No. 6265550
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Puryi
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/270.117
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/878,563
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0323 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:

LIBRARY: GenBank
CLONE: 1203820
US-09-270-117-3

Query*Match 1.4%; Score 7; DB 3; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 KAAQPKS 35
|||||
Db 318 KAAQPKS 324

RESULT 35
US-08-878-563A-1
Sequence 1, Application US/08878563A
Patent No. 5891674
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Puryi
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,563A
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0323 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 534 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT04
CLONE: 918158
US-08-878-563A-1

Query*Match 1.4%; Score 7; DB 2; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 KAAQPKS 35
|||||
Db 317 KAAQPKS 323

RESULT 36
US-09-270-117-1
Sequence 1, Application US/09270117
Patent No. 6265550
GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purni
TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/270,117
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/878,563
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0323 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 534 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT04
CLONE: 918158
US-09-270-117-1

Query Match 1.4%; Score 7; DB 3; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 KAAOPKS 35
Db 317 KAAOPKS 323

RESULT 37
US-09-328-352-5771
Sequence 5771, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5771
LENGTH: 539
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5771

Query Match 1.4%; Score 7; DB 4; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 GENALKA 263

Db 519 GENALKA 525

RESULT 38
US-09-252-991A-20407
Sequence 20407, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20407
LENGTH: 580
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20407

Query Match 1.4%; Score 7; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 PEGAPQA 376
Db 466 PEGAPQA 472

RESULT 39
US-09-252-991A-19164
Sequence 19164, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19164
LENGTH: 616
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19164

Query Match 1.4%; Score 7; DB 4; Length 616;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 RLADGVE 102
Db 432 RLADGVE 438

RESULT 40
US-09-252-991A-26100
Sequence 26100, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26100
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26100

Query Match 1.4%; Score 7; DB 4; Length 727;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 96 RLADGVE 102
|||
Db 637 RLADGVE 643

Search completed: August 27, 2003, 18:46:11
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:45:18 ; Search time 60 seconds

(without alignments)
1144.408 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502
Sequence: 1 MSKPTLKITTLICALSALML.....NKGGQISADVAKKTKPEN 502Scoring table: OLIGO
Gapox 60.0 , Gapext 60.0

Searched: 510680 seqs, 136781880 residues

Word size : 0

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.4	39	12	US-09-890-688-140
2	7	1.4	71	9	US-09-864-761-44403
3	7	1.4	104	15	US-10-026-741-100
4	7	1.4	156	15	US-10-101-464A-509
5	7	1.4	157	15	US-10-156-761-13774
6	7	1.4	180	14	US-10-027-450-28
7	7	1.4	233	14	US-10-138-787-7
8	7	1.4	238	9	US-09-904-954-2
9	7	1.4	238	12	US-10-241-220-72
10	7	1.4	278	10	US-09-738-626-5155
11	7	1.4	300	15	US-10-156-761-14202
12	7	1.4	312	10	US-09-967-347-11
13	7	1.4	411	12	US-10-092-947A-37
14	7	1.4	436	9	US-09-730-772-14
15	7	1.4	436	9	US-09-735-849-14

16	7	1.4	438	15	US-10-062-848-2	Sequence 2, Appli
17	7	1.4	439	10	US-09-925-637-20	Sequence 20, Appli
18	7	1.4	439	15	US-10-084-205-20	Sequence 20, Appli
19	7	1.4	450	15	US-10-062-848-15	Sequence 15, Appli
20	7	1.4	451	11	US-09-843-250-24	Sequence 24, Appli
21	7	1.4	518	15	US-10-278-481-18	Sequence 18, Appli
22	7	1.4	519	15	US-10-211-239-9	Sequence 1, Appli
23	7	1.4	549	15	US-10-156-761-14029	Sequence 14029, A
24	7	1.4	611	9	US-09-815-242-12111	Sequence 12111, A
25	7	1.4	724	15	US-10-156-761-9578	Sequence 9578, Ap
26	7	1.4	792	10	US-09-995-587A-11	Sequence 11, Appli
27	7	1.4	844	12	US-09-882-227-308	Sequence 308, App
28	7	1.4	859	8	US-08-945-749-1	Sequence 1, Appli
29	7	1.4	886	8	US-08-781-986A-5235	Sequence 5235, Ap
30	7	1.4	960	15	US-10-188-308-21	Sequence 21, Appli
31	7	1.4	960	15	US-10-188-296-21	Sequence 21, Appli
32	7	1.4	960	15	US-10-188-341-21	Sequence 21, Appli
33	7	1.4	960	15	US-10-188-297-21	Sequence 21, Appli
34	7	1.4	962	15	US-10-188-308-3	Sequence 3, Appli
35	7	1.4	962	15	US-10-188-308-24	Sequence 24, Appli
36	7	1.4	962	15	US-10-188-296-3	Sequence 3, Appli
37	7	1.4	962	15	US-10-188-296-24	Sequence 24, Appli
38	7	1.4	962	15	US-10-188-341-3	Sequence 3, Appli
39	7	1.4	962	15	US-10-188-341-24	Sequence 24, Appli
40	7	1.4	962	15	US-10-188-297-3	Sequence 3, Appli
41	7	1.4	962	15	US-10-188-297-24	Sequence 24, Appli
42	7	1.4	962	15	US-10-174-613-4	Sequence 4, Appli
43	7	1.4	962	15	US-10-174-613-5	Sequence 5, Appli
44	7	1.4	987	15	US-10-188-308-22	Sequence 22, Appli
45	7	1.4	987	15	US-10-188-296-22	Sequence 22, Appli
46	7	1.4	987	15	US-10-188-341-22	Sequence 22, Appli
47	7	1.4	987	15	US-10-188-297-22	Sequence 22, Appli
48	7	1.4	988	11	US-09-875-321-5	Sequence 5, Appli
49	7	1.4	988	11	US-09-875-321-12	Sequence 12, Appli
50	7	1.4	988	15	US-10-128-323-2	Sequence 2, Appli
51	7	1.4	988	15	US-10-162-012-5	Sequence 5, Appli
52	7	1.4	988	15	US-10-162-012-12	Sequence 12, Appli
53	7	1.4	988	15	US-10-174-613-2	Sequence 2, Appli
54	7	1.4	988	15	US-10-174-613-3	Sequence 3, Appli
55	7	1.4	989	15	US-10-188-308-4	Sequence 4, Appli
56	7	1.4	989	15	US-10-188-308-23	Sequence 23, Appli
57	7	1.4	989	15	US-10-188-296-4	Sequence 4, Appli
58	7	1.4	989	15	US-10-188-296-23	Sequence 23, Appli
59	7	1.4	989	15	US-10-188-341-4	Sequence 4, Appli
60	7	1.4	989	15	US-10-188-341-23	Sequence 23, Appli
61	7	1.4	989	15	US-10-188-297-23	Sequence 23, Appli
62	7	1.4	989	15	US-10-188-297-23	Sequence 23, Appli
63	7	1.4	989	15	US-10-174-613-6	Sequence 6, Appli
64	7	1.4	1095	15	US-10-128-714-3039	Sequence 3039, Ap
65	7	1.4	1174	15	US-10-174-613-7	Sequence 7, Appli
66	7	1.4	1277	15	US-10-128-714-8039	Sequence 8039, Ap
67	7	1.4	1676	15	US-10-128-714-8246	Sequence 8246, Ap
68	7	1.2	11	15	US-10-108-795-28	Sequence 28, Appli
69	6	1.2	11	15	US-10-108-795-29	Sequence 29, Appli
70	6	1.2	11	15	US-10-115-365-28	Sequence 28, Appli
71	6	1.2	11	15	US-10-115-365-29	Sequence 29, Appli
72	6	1.2	23	11	US-09-774-639-331	Sequence 331, App
73	6	1.2	23	11	US-09-969-720-263	Sequence 263, App
74	6	1.2	25	9	US-09-864-761-39462	Sequence 39462, A
75	6	1.2	31	9	US-09-880-578-36	Sequence 36, Appli
76	6	1.2	34	9	US-09-864-761-39670	Sequence 39670, A
77	6	1.2	34	9	US-09-864-761-42206	Sequence 42206, A
78	6	1.2	38	10	US-09-919-473-12	Sequence 12, Appli
79	6	1.2	38	15	US-10-099-766-6	Sequence 6, Appli
80	6	1.2	40	9	US-09-864-761-43181	Sequence 43181, A
81	6	1.2	40	9	US-09-682-706-15	Sequence 15, Appli
82	6	1.2	47	8	US-08-424-550B-556	Sequence 556, App
83	6	1.2	49	9	US-09-864-761-36011	Sequence 36011, A
84	6	1.2	57	10	US-09-738-626-5498	Sequence 5498, Ap
85	6	1.2	58	11	US-09-764-891-4535	Sequence 4535, Ap
86	6	1.2	59	9	US-09-864-761-41409	Sequence 41409, A
87	6	1.2	59	9	US-09-864-761-46850	Sequence 46850, A
88	6	1.2	60	15	US-10-179-784-8	Sequence 8, Appli

6 1.2 386 15 US-10-034-937-2 Sequence 2, Appl
6 1.2 386 15 US-10-122-706-2 Sequence 2, Appl
6 1.2 387 9 US-09-866-987-9 Sequence 9, Appl
6 1.2 387 15 US-10-034-937-42 Sequence 42, Appl
6 1.2 387 15 US-10-156-761-14269 Sequence 14269, A
6 1.2 387 15 US-10-156-761-14583 Sequence 14583, A
6 1.2 388 15 US-09-880-578-17 Sequence 17, Appl
6 1.2 388 15 US-10-161-572-47 Sequence 47, Appl
6 1.2 388 15 US-10-230-318-1 Sequence 1, Appl
6 1.2 389 9 US-09-880-578-22 Sequence 22, Appl
6 1.2 389 9 US-09-880-578-24 Sequence 24, Appl
6 1.2 389 9 US-09-880-578-25 Sequence 25, Appl
6 1.2 389 9 US-09-880-578-26 Sequence 26, Appl
6 1.2 389 9 US-09-880-578-27 Sequence 27, Appl
6 1.2 389 9 US-09-880-578-28 Sequence 28, Appl
6 1.2 389 9 US-09-880-578-29 Sequence 29, Appl
6 1.2 389 9 US-09-880-578-30 Sequence 30, Appl
6 1.2 389 9 US-09-880-578-31 Sequence 31, Appl
6 1.2 390 10 US-09-939-521-10 Sequence 10, Appl
6 1.2 390 15 US-10-034-937-4 Sequence 4, Appl
6 1.2 390 15 US-10-034-937-6 Sequence 6, Appl
6 1.2 390 15 US-10-034-937-8 Sequence 8, Appl
6 1.2 390 15 US-10-034-937-10 Sequence 10, Appl
6 1.2 390 15 US-10-034-937-12 Sequence 12, Appl
6 1.2 390 15 US-10-034-937-14 Sequence 14, Appl
6 1.2 390 15 US-10-034-937-16 Sequence 16, Appl
6 1.2 390 15 US-10-034-937-18 Sequence 18, Appl
6 1.2 390 15 US-10-034-937-20 Sequence 20, Appl
6 1.2 390 15 US-10-034-937-22 Sequence 22, Appl
6 1.2 390 15 US-10-034-937-24 Sequence 24, Appl
6 1.2 390 15 US-10-034-937-26 Sequence 26, Appl
6 1.2 390 15 US-10-034-937-28 Sequence 28, Appl
6 1.2 390 15 US-10-034-937-30 Sequence 30, Appl
6 1.2 390 15 US-10-034-937-32 Sequence 32, Appl
6 1.2 390 15 US-10-034-937-34 Sequence 34, Appl
6 1.2 390 15 US-10-034-937-36 Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-890-688-140
; Sequence 140, Application US/09890688
; Publication No. US20030144475A1
; GENERAL INFORMATION:
; APPLICANT: Seishi KATO
; APPLICANT: Chikashi EGUCHI
; APPLICANT: Mihoro SAEKI
; TITLE OF INVENTION: Human Proteins and cDNAs thereof
; FILE REFERENCE: 2001-1102A/WMC/00653
; CURRENT APPLICATION NUMBER: US/09/890, 688
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: JP 11-346863
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 11-34684
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: JP 2000-31062
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: JP 2000-34091
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-34090
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: JP 2000-35829
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-35839
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: JP 2000-71161
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: JP 2000-160851
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 140
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-890-688-140
Query Match 1.4%; Score 7; DB 12; Length 39;
Best local similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 150 GCGAENS 156
Db 6 GCGAENS 12
RESULT 2
US-09-864-761-44403
; Sequence 44403, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180, 312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207, 456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632, 366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236, 359.
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234, 687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608, 408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774, 203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 44403
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005015.2

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.78
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.78
OTHER INFORMATION: EST HUMAN HIT: AA077633.1, EVALUE 2.00e-19
OTHER INFORMATION: SWISSPROT HIT: O42184, EVALUE 3.00e-11
US-09-864-761-44403

Query Match 1.4%; Score 7; DB 9; Length 71;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 ARAKTK 500
DB 28 ARAKTK 34

RESULT 3
US-10-026-741-100
Sequence 100, Application US/10026741
Publication No. US20030049604A1
GENERAL INFORMATION:
APPLICANT: CHARNEAU, PIERRE
CLAVEL, FRANCOISE
BORMAN, ANDREW
OUILLENT, CAROLINE
GUETARD, DENISE
MONTAGNIER, LUC
DOMJON DE SAINT-MARTIN, JACQUELINE
COHEN, JACQUES

TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR
SUBTYPE) ANTIGENS

NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC

COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/026,741
FILING DATE: 27-Dec-2001
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/817,441
FILING DATE: 31-AUG-1998

APPLICATION NUMBER: PCT/FR 95/01391
FILING DATE: 20-OCT-1995

APPLICATION NUMBER: FR 9412554
FILING DATE: 20-OCT-1994

APPLICATION NUMBER: FR 9502526
FILING DATE: 03-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03260.6005-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-408-4000

TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 100:

SEQUENCE CHARACTERISTICS:

LENGTH: 104 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 100:
US-10-026-741-100

Query Match 1.4%; Score 7; DB 15; Length 104;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DQDIANV 473
DB 83 DQDIANV 89

RESULT 4
US-10-101-464A-509
Sequence 509, Application US/10101464A
Publication No. US20030046728A1
GENERAL INFORMATION:

APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Nicolaas

APPLICANT: Higgins, Colleen M.

TITLE OF INVENTION: Compositions Isolated from Plant Cells

TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling

FILE REFERENCE: 11000.1020c2

CURRENT APPLICATION NUMBER: US/10/101,464A

CURRENT FILING DATE: 2002-03-18

PRIOR APPLICATION NUMBER: 09/704,302

PRIOR FILING DATE: 2000-11-01

PRIOR APPLICATION NUMBER: 09/228,986

PRIOR FILING DATE: 1999-01-12

PRIOR APPLICATION NUMBER: 60/162,866

PRIOR FILING DATE: 1999-11-01

PRIOR APPLICATION NUMBER: PCT/US00/00724

PRIOR FILING DATE: 2000-01-11

NUMBER OF SEQ ID NOS: 989

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 509

TYPE: PRT

LENGTH: 156

ORGANISM: Eucalyptus grandis

US-10-101-464A-509

Query Match 1.4%; Score 7; DB 15; Length 156;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 NGISGKI 449
DB 115 NGISGKI 121

RESULT 5
US-10-156-761-13774
Sequence 13774, Application US/10156761

Publication No. US20030119018A1
GENERAL INFORMATION:

APPLICANT: OKURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 13774

LENGTH: 157
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13774

Query Match
Best Local Similarity 1.4%; Score 7; DB 15; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 HOPDGK 416
Db 140 HOPDGK 146

RESULT 6
US-10-027-450-28
Sequence 28, Application US/10027450
Publication No. US20020102715A1
GENERAL INFORMATION:
APPLICANT: Falco, Saverio Carl
APPLICANT: Hitz, William D.
APPLICANT: Kinney, Anthony J.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Rafaleki, J. Antoni
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
FILE REFERENCE: BB-1126
CURRENT APPLICATION NUMBER: US/10/027,450
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/063,423
PRIOR FILING DATE: 1997 October 28
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 28
LENGTH: 180
TYPE: PRT
ORGANISM: Triticum aestivum
US-10-027-450-28

Query Match
Best Local Similarity 1.4%; Score 7; DB 14; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 KVEGEE 355
Db 58 KVEGEE 64

RESULT 7
US-10-138-787-7
Sequence 7, Application US/10138787
Publication No. US20020172984A1
GENERAL INFORMATION:
APPLICANT: Holland, Sacha
APPLICANT: Mbemali, Geraldine
APPLICANT: Pawsen, Tony
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
TITLE OF INVENTION: TYROSINE KINASES
FILE REFERENCE: 11757,23USWO
CURRENT APPLICATION NUMBER: US/10/138,787
PRIOR FILING DATE: 2002-05-03
PRIOR APPLICATION NUMBER: US/09/214,631
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: PCT/CA97/00473
PRIOR FILING DATE: 1997-07-04
PRIOR APPLICATION NUMBER: 60/021,272
PRIOR FILING DATE: 1996-07-05
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 233
TYPE: PRT
ORGANISM: Homo sapiens

US-10-138-787-7

Query Match
Best Local Similarity 1.4%; Score 7; DB 14; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
Db 74 GPGGAE 80

RESULT 8
US-09-904-954-2
Sequence 2, Application US/09904954
Patent No. US20020010325A1
GENERAL INFORMATION:
APPLICANT: BECKMANN, M. P.
CERRETTI, DOUGLAS P.
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
RECEPTOR HERK
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNEX CORPORATION
STREET: 51 UNIVERSITY STREET
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/904,954
FILING DATE: 12-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/240,124
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/114,426
FILING DATE: 30-AUG-1993
APPLICATION NUMBER: US 08/109,745
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: SEESE, KATHRYN A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2814-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-904-954-2

Query Match
Best Local Similarity 1.4%; Score 7; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
Db 78 GPGGAE 84

RESULT 9
US-10-241-220-72
Sequence 72, Application US/10241220

Publication No. US20030148408A1
GENERAL INFORMATION:
APPLICANT: Frantz, Gretchen
APPLICANT: Hillan, Kenneth J.
APPLICANT: Phillips, Heidi
APPLICANT: Polakis, Paul
APPLICANT: Spencer, Susan
APPLICANT: Williams, P. Mickey
APPLICANT: Wu, Thomas
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
FILE REFERENCE: P5010R1-US
CURRENT APPLICATION NUMBER: US/10/241,220
CURRENT FILING DATE: 2002-12-13
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 72
LENGTH: 238
TYPE: PRT
ORGANISM: Homo Sapien
US-10-241-220-72

Query Match 1.4%; Score 7; DB 12; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAAE 154
|||||
Db 78 GPGGAAE 84

RESULT 10
US-09-738-626-5155
Sequence 5155, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5155
LENGTH: 278
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5155

Query Match 1.4%; Score 7; DB 10; Length 278;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 RLADGVE 102
|||||
Db 262 RLADGVE 268

RESULT 11
US-10-156-761-14202
Sequence 14202, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14202
LENGTH: 300
TYPE: PRT
ORGANISM: Streptomyces avermectilis
US-10-156-761-14202

Query Match 1.4%; Score 7; DB 15; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 RLADGVE 102
|||||
Db 266 RLADGVE 272

RESULT 12
US-09-967-347-11
Sequence 11, Application US/09967347
Patent No. US20020115829A1
GENERAL INFORMATION:
APPLICANT: Finlay, B. Brett
APPLICANT: Kenny, Brendan
APPLICANT: Stein, Markus
APPLICANT: Domeneberg, Michael S.
TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN EspA
FILE REFERENCE: 482112.401D1
CURRENT APPLICATION NUMBER: US/09/967,347
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: 09/171,517
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: PCT/CA97/00265
PRIOR FILING DATE: 1997-04-23
PRIOR APPLICATION NUMBER: 60/015,999
PRIOR FILING DATE: 1996-04-23
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 312
TYPE: PRT
ORGANISM: Escherichia coli
US-09-967-347-11

Query Match 1.4%; Score 7; DB 10; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 ADDVAKA 496
|||||
Db 187 ADDVAKA 193

RESULT 13
US-10-092-947A-37
; Sequence 37, Application US/10092947A
; Publication NO. US20030134353A1
; GENERAL INFORMATION:
; APPLICANT: WOLFE, Anne M
; APPLICANT: APPEL, Karen F
; APPLICANT: PETERSEN, Jesper F
; APPLICANT: POULSEN, Ulla
; APPLICANT: ARNAU, Jose
; APPLICANT: JACOBSEN, Mette D
; TITLE OF INVENTION: MUCOR RECOMBINANT GENE EXPRESSION
; FILE REFERENCE: WOLFE=3
; CURRENT APPLICATION NUMBER: US/10/092, 947A
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US 60/274,650
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-092-947A-37

Query Match 1.4%; Score 7; DB 12; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 RAASIVA 442
|||||
Db 368 RAASIVA 374

RESULT 14
US-09-730-772-14
; Sequence 14, Application US/09730772
; Patent No. US2001001131A1
; GENERAL INFORMATION:
; APPLICANT: Luyten, Frank P.
; APPLICANT: Moos, Jr., Malcolm
; APPLICANT: Chang, Steven Chao-Huan
; TITLE OF INVENTION: CARTILAGE-DERIVED MORPHOGENETIC
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/730, 772
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,081
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bartfeld, Neil S
; REGISTRATION NUMBER: 39,901
; REFERENCE/DOCKET NUMBER: NIH099,001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-730-772-14

Query Match 1.4%; Score 7; DB 9; Length 436;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
|||||
Db 279 GPGGAE 285

RESULT 15
US-09-735-849-14
; Sequence 14, Application US/09735849
; Patent No. US20010037017A1
; GENERAL INFORMATION:
; APPLICANT: Luyten, Frank P.
; APPLICANT: Moos, Jr., Malcolm
; APPLICANT: Chang, Steven Chao-Huan
; TITLE OF INVENTION: CARTILAGE-DERIVED MORPHOGENETIC
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/735, 849
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,081
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bartfeld, Neil S
; REGISTRATION NUMBER: 39,901
; REFERENCE/DOCKET NUMBER: NIH099,001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX:
; INFORMATION FOR 'SEQ ID NO: 14':
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-735-849-14

Query Match 1.4%; Score 7; DB 9; Length 436;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
|||||
Db 279 GPGGAE 285

RESULT 16
US-10-062-848-2
; Sequence 2, Application US/10062848
; Publication No. US20030092155A1
; GENERAL INFORMATION:
; APPLICANT: KOSTREMA, Dirk
; APPLICANT: PASAMONTES, Luis
; APPLICANT: TOMSCHY, Andrea
; APPLICANT: van LOON, Adolphus
; APPLICANT: VOGEL, Kurt
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: MODIFIED PHYTASES
; FILE REFERENCE: Modified Phytases
; CURRENT APPLICATION NUMBER: US/10/062,848
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/044,718
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: EP 97810175.6
; PRIOR FILING DATE: 1997-03-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Aspergillus terreus
US-10-062-848-2

Query Match 1.4%; Score 7; DB 15; Length 438;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 QLSADVV 493
DB 223 QLSADVV 229

RESULT 17
US-09-925-637-20
; Sequence 20, Application US/09925637
; Patent No. US2002010338A1
; GENERAL INFORMATION:
; APPLICANT: Choi
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Polypeptides
; FILE REFERENCE: PB560
; CURRENT APPLICATION NUMBER: US/09/925,637
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/23773
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 60/151,933
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 08/781,986
; PRIOR FILING DATE: 1997-01-03
; PRIOR APPLICATION NUMBER: US 08/956,171
; PRIOR FILING DATE: 1997-10-20
; PRIOR APPLICATION NUMBER: US 60/009,861
; PRIOR FILING DATE: 1996-01-06
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-637-20

Query Match 1.4%; Score 7; DB 10; Length 439;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 RAASIVA 442
DB 301 RAASIVA 307

RESULT 18
US-10-084-205-20
; Sequence 20, Application US/10084205
; Publication No. US20030049648A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil
; TITLE OF INVENTION: 37 Staphylococcus aureus Genes and Polypeptides
; FILE REFERENCE: PB515P1
; CURRENT APPLICATION NUMBER: US/10/084,205
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: PCT/US00/23773
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/151,933
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 20
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-084-205-20

Query Match 1.4%; Score 7; DB 15; Length 439;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 RAASIVA 442
DB 301 RAASIVA 307

RESULT 19
US-10-062-848-15
; Sequence 15, Application US/10062848
; Publication No. US20030092155A1
; GENERAL INFORMATION:
; APPLICANT: KOSTREMA, Dirk
; APPLICANT: PASAMONTES, Luis
; APPLICANT: TOMSCHY, Andrea
; APPLICANT: van LOON, Adolphus
; APPLICANT: VOGEL, Kurt
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: MODIFIED PHYTASES
; FILE REFERENCE: Modified Phytases
; CURRENT APPLICATION NUMBER: US/10/062,848
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/044,718
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: EP 97810175.6
; PRIOR FILING DATE: 1997-03-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Aspergillus terreus
US-10-062-848-15

Query Match 1.4%; Score 7; DB 15; Length 450;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 QLSADVV 493
DB 235 QLSADVV 241

RESULT 20
US-09-843-250-24
; Sequence 24, Application US/09843250
; Publication No. US2003002235A1
; GENERAL INFORMATION:
; APPLICANT: Parales, R.

APPLICANT: Gibson, D.
APPLICANT: Resnick, S.
APPLICANT: Lee, K.
TITLE OF INVENTION: No. US20030022335A1el naphthalene dioxigenase and methods for the
FILE REFERENCE: 875.006US2
CURRENT APPLICATION NUMBER: US/09/843.250
CURRENT FILING DATE: 2001-04-26
PRIOR APPLICATION NUMBER: PCT/US99/25079
PRIOR FILING DATE: 1999-10-26
PRIOR APPLICATION NUMBER: US 60/105,575
PRIOR FILING DATE: 1998-10-26
NUMBER OF SEQ ID NOS: 65
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 451
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: A polypeptide encoded by SEQ ID NO:13.
US-09-843-250-24

Query Match 1.4%; Score 7; DB 11; Length 451;
Best Local Similarity 100.0%; Pred. No. 3.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 PAKVVK 87
Db 189 PAKVVK 195

RESULT 21
US-10-278-481-18
Sequence 18, Application US/10278481
Publication No. US20030113803A1
GENERAL INFORMATION:
APPLICANT: KARO BIO AB
TITLE OF INVENTION: Ophan receptor
NUMBER OF SEQUENCES: 19
STREET: c/o Center for Biotechnology and Department
of Medical Nutrition, Karolinska Nutrition, Karolinska
Institute
CITY: Huddinge
COUNTRY: Sweden
ZIP: S-14186
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/278.481
FILING DATE: 23-Oct-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333.057
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/836,620
FILING DATE: <Unknown>
APPLICATION NUMBER: GB 9518272.1
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: GB 9605550.4
FILING DATE: 15-MAR-1996
APPLICATION NUMBER: GB 9607532.0
FILING DATE: 11-APR-1996
APPLICATION NUMBER: GB 9609576.5
FILING DATE: 08-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: GUSTAFSSON, Jan-Ake
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 amino acids
TYPE: amino acid
TOPOLOGY: linear
ORIGINAL SOURCE:

ORGANISM: Homo sapiens
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-278-481-18

Query Match 1.4%; Score 7; DB 15; Length 518;
Best Local Similarity 100.0%; Pred. No. 3.5e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
Db 466 GPGGAE 472

RESULT 22
US-10-211-239-1
Sequence 1, Application US/10211239
Publication No. US20030103965A1
GENERAL INFORMATION:
APPLICANT: Jung, Birgit
APPLICANT: Kraut, No. US20030103965A1bert
APPLICANT: Mueller, Stefan
TITLE OF INVENTION: Method for identifying Substances which Positively
Influence Inflammatory Conditions
FILE REFERENCE: 0652.2340001
CURRENT APPLICATION NUMBER: US/10/211,239
CURRENT FILING DATE: 2002-08-05
PRIOR APPLICATION NUMBER: US 60/315,775
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: EP 01119003.0
PRIOR FILING DATE: 2001-08-06
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 519
TYPE: PRT
ORGANISM: Homo Sapiens
US-10-211-239-1

Query Match 1.4%; Score 7; DB 15; Length 519;
Best Local Similarity 100.0%; Pred. No. 3.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 GPGGAE 154
Db 464 GPGGAE 470

RESULT 23
US-10-156-761-14029
Sequence 14029, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OKURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIRAMA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14029
LENGTH: 549
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14029

Query Match 1.4%; Score 7; DB 15; Length 549;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 VMLADG 100
|||||
Db 326 VMLADG 332

RESULT 24
US-09-815-242-12111
; Sequence 12111, Application US/09815242

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12111
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12111

Query Match 1.4%; Score 7; DB 9; Length 611;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 RLADGVE 102
|||||
Db 427 RLADGVE 433

RESULT 25
US-10-156-761-9578
; Sequence 9578, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9578
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9578

Query Match 1.4%; Score 7; DB 15; Length 724;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 LADGVEY 103
|||||
Db 260 LADGVEY 266

RESULT 26
US-09-995-587A-11
; Sequence 11, Application US/0995587A
; Patent No. US20020127681A1
; GENERAL INFORMATION:
; APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO
; APPLICANT: VAN GEEL-SCHUTTEN, GERTRIDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES
; FILE REFERENCE: B043667-CIP
; CURRENT APPLICATION NUMBER: US/09/995,587A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 09/604,958
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201872.9
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 792
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri

; NAME/KEY: MOD RES
; LOCATION: (495)..(496)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (737)
; OTHER INFORMATION: Thr or Pro
US-09-995-587A-11

Query Match 1.4%; Score 7; DB 10; Length 792;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GALGILK 349
|||||
Db 461 GALGILK 467

RESULT 27
US-09-882-227-308
; Sequence 308, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleenhouw, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois

APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20030158396a1 Helicobacter Polypeptides in the
TITLE OF INVENTION: Genome
FILE REFERENCE: 06132/047002
CURRENT APPLICATION NUMBER: US/09/882,227
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/902,615
PRIOR FILING DATE: 1997-07-29
NUMBER OF SEQ ID NOS: 638
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 308
LENGTH: 844
TYPE: PRT
ORGANISM: Helicobacter pylori.
US-09-882-227-308

Query Match 1.4%; Score 7; DB 12; Length 844;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 EYSHKQ 363
DB 553 EYSHKQ 559

RESULT 28
US-08-945-749-1
Sequence 1, Application US/08945749
Publication No. US20020138880A1
GENERAL INFORMATION:
APPLICANT: GARDNER, Richard C
APPLICANT: MACDIARMID, Colin W
APPLICANT: HAY, Robert J
APPLICANT: Auckland Uniservices Limited
APPLICANT: New Zealand Pastoral Agriculture Research Institute
TITLE OF INVENTION: ALUMINUM RESISTANCE GENE
FILE REFERENCE: 08/945,749
CURRENT APPLICATION NUMBER: US/08/945,749
CURRENT FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: PCT/NZ96/0035
EARLIER FILING DATE: 1996-05-01
EARLIER APPLICATION NUMBER: NZ 272039
EARLIER FILING DATE: 1995-05-01
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 859
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-08-945-749-1

Query Match 1.4%; Score 7; DB 8; Length 859;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 KVEEEN 355
DB 388 KVEEEN 394

RESULT 29
US-08-781-986A-5235
Sequence 5235, Application US/08781986A
Publication No. US2003005436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville

STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 5235:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-781-986A-5235

Query Match 1.4%; Score 7; DB 8; Length 886;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 AOPKST 37
DB 295 AOPKST 301

RESULT 30
US-10-188-308-21
Sequence 21, Application US/10188308
Publication No. US20030077735A1
GENERAL INFORMATION:
APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
APPLICANT: STUMER, WALTER
APPLICANT: BECKH, SYMNOVE
APPLICANT: BRUGEMANN, ANDREA
APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
APPLICANT: PEREZ, ARACELI SANCHEZ
APPLICANT: WESELOH, RUDIGER
TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC
TITLE OF INVENTION: APPLICATIONS
TITLE OF INVENTION: THEROPF
FILE REFERENCE: MPG-8
CURRENT APPLICATION NUMBER: US/10/188,308
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: US/09/694,777
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: PCT/EP99/02695
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: EP 98 10 7268.9
PRIOR FILING DATE: 1998-04-21
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 960
TYPE: PRT
ORGANISM: Bovine sp.
US-10-188-308-21

Query Match 1.4%; Score 7; DB 15; Length 960;

Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386
|||||
DB 200 QCAPKTP 206

RESULT 31
US-10-188-296-21

; Sequence 21, Application US/10188296
; Publication No. US20030087377A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYMNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,296

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: US/09/694,777

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: PCT/EP99/02695

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: EP 98 10 7268.9

; PRIOR FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21

; LENGTH: 960

; TYPE: PRT

; ORGANISM: Bovine sp.

US-10-188-296-21

Query Match 1.4%; Score 7; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386
|||||
DB 200 QCAPKTP 206

RESULT 32
US-10-188-341-21

; Sequence 21, Application US/10188341

; Publication No. US20030087378A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYMNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,341

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: US/09/694,777

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: PCT/EP99/02695

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: EP 98 10 7268.9

; PRIOR FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Bovine sp.
US-10-188-341-21

Query Match 1.4%; Score 7; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386
|||||
DB 200 QCAPKTP 206

RESULT 33
US-10-188-297-21

; Sequence 21, Application US/10188297

; Publication No. US20030092120A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYMNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,297

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: US/09/694,777

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: PCT/EP99/02695

; PRIOR FILING DATE: 1999-04-21

; PRIOR APPLICATION NUMBER: EP 98 10 7268.9

; PRIOR FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21

; LENGTH: 960

; TYPE: PRT

; ORGANISM: Bovine sp.

US-10-188-297-21

Query Match 1.4%; Score 7; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 QCAPKTP 386
|||||
DB 200 QCAPKTP 206

RESULT 34
US-10-188-308-3

; Sequence 3, Application US/10188308

; Publication No. US2003007735A1

; GENERAL INFORMATION:

; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL

; APPLICANT: STUMMER, WALTER

; APPLICANT: BECKH, SYMNOVE

; APPLICANT: BRUGEMANN, ANDREA

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO

; APPLICANT: PEREZ, ARACELI SANCHEZ

; APPLICANT: WESELOH, RUDIGER

; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC

; TITLE OF INVENTION: APPLICATIONS

; FILE REFERENCE: MPG-8

; CURRENT APPLICATION NUMBER: US/10/188,308

; CURRENT FILING DATE: 2002-07-01

```

; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-308-3
```

```

Query Match      1.4% Score 7; DB 15; Length 962;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      380 QCAPKTP 386
        |||||
Db      200 QCAPKTP 206
```

```

RESULT 35
US-10-188-308-24
; Sequence 24, Application US/10188308
; Publication No. US20030077735A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUMMER, WALTER
; APPLICANT: BECKH, SYNNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC
; TITLE OF INVENTION: APPLICATIONS
; FILE REFERENCE: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,308
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-188-308-24
```

```

Query Match      1.4% Score 7; DB 15; Length 962;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      380 QCAPKTP 386
        |||||
Db      200 QCAPKTP 206
```

```

RESULT 36
US-10-188-296-3
; Sequence 3, Application US/10188296
; Publication No. US20030087377A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUMMER, WALTER
; APPLICANT: BECKH, SYNNOVE
; APPLICANT: BRUGEMANN, ANDREA
```

```

; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,296
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-296-3
```

```

Query Match      1.4% Score 7; DB 15; Length 962;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      380 QCAPKTP 386
        |||||
Db      200 QCAPKTP 206
```

```

RESULT 37
US-10-188-296-24
; Sequence 24, Application US/10188296
; Publication No. US20030087377A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUMMER, WALTER
; APPLICANT: BECKH, SYNNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,296
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-188-296-24
```

```

Query Match      1.4% Score 7; DB 15; Length 962;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      380 QCAPKTP 386
        |||||
Db      200 QCAPKTP 206
```

```

RESULT 38
US-10-188-341-3
```

```
; Sequence 3, Application US/10188341
; Publication No. US20030087378A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUHRER, WALTER
; APPLICANT: BECKH, SYMNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,341
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-341-3
```

```
Query Match 1.4%; Score 7; DB 15; Length 962;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 380 OEAPKTP 386
Db 200 OEAPKTP 206
```

```
RESULT 39
US-10-188-341-24
; Sequence 24, Application US/10188341
; Publication No. US20030087378A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUHRER, WALTER
; APPLICANT: BECKH, SYMNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,341
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-188-341-24
```

```
Query Match 1.4%; Score 7; DB 15; Length 962;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 380 OEAPKTP 386
Db 200 OEAPKTP 206
```

```
RESULT 40
US-10-188-297-3
; Sequence 3, Application US/10188297
; Publication No. US20030092120A1
; GENERAL INFORMATION:
; APPLICANT: PARDO-FERNANDEZ, LUIS ANGEL
; APPLICANT: STUHRER, WALTER
; APPLICANT: BECKH, SYMNOVE
; APPLICANT: BRUGEMANN, ANDREA
; APPLICANT: FERNANDEZ-MIRANDA, DONATO DEL CAMINO
; APPLICANT: PEREZ, ARACELI SANCHEZ
; APPLICANT: WESELOH, RUDIGER
; TITLE OF INVENTION: NOVEL HUMAN K+ ION CHANNEL AND THERAPEUTIC APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MPG-8
; CURRENT APPLICATION NUMBER: US/10/188,297
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/694,777
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: PCT/EP99/02695
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: EP 98 10 7268.9
; PRIOR FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-297-3
```

```
Query Match 1.4%; Score 7; DB 15; Length 962;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 380 OEAPKTP 386
Db 200 OEAPKTP 206
```

```
Search completed: August 27, 2003, 18:54:16
Job time : 72 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 27, 2003, 18:32:47 ; Search time 86 Seconds

(without alignments)
926.520 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 502
Sequence: 1 MSKPLTKITTLICALSALML.....NKGQISADVAKAKTKPEN 502

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A.Geneseq.19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	502	100.0	502	22	AAB67668
2	502	100.0	502	22	AAB67669
3	197	39.2	197	23	AAO17561
4	34	6.8	392	24	ABP77627
5	8	1.6	10	20	AAO5975
6	8	1.6	415	22	ABG18127
7	8	1.6	497	21	AAI74923
8	8	1.6	497	21	AAI74924
9	8	1.6	497	21	AAI74925

10	8	1.6	508	24	ABP80199
11	8	1.6	697	21	AAV93290
12	8	1.6	697	21	AAV93303
13	8	1.6	995	23	ABP65660
14	7	1.4	33	21	AAI12246
15	7	1.4	39	22	AAI93319
16	7	1.4	71	22	ABG55536
17	7	1.4	71	22	ABG40293
18	7	1.4	71	22	AAI61086
19	7	1.4	71	22	AAI73789
20	7	1.4	71	22	AAI33976
21	7	1.4	71	22	ABG43675
22	7	1.4	90	22	ABG60712
23	7	1.4	94	22	AAO10927
24	7	1.4	104	17	AAI07245
25	7	1.4	105	22	ABG26287
26	7	1.4	111	16	AAI72394
27	7	1.4	111	16	AAI72394
28	7	1.4	118	21	AAI27869
29	7	1.4	124	22	ABG08519
30	7	1.4	124	22	AAO01520
31	7	1.4	130	23	ABG55608
32	7	1.4	144	21	AAI07374
33	7	1.4	149	21	AAI07373
34	7	1.4	156	21	AAI25190
35	7	1.4	167	22	ABG58068
36	7	1.4	172	23	ABP38126
37	7	1.4	180	21	AAI28428
38	7	1.4	180	22	AAI86749
39	7	1.4	185	22	ABP71959
40	7	1.4	187	22	ABP71208
41	7	1.4	189	22	AAI62704
42	7	1.4	224	19	AAI69164
43	7	1.4	228	22	ABG62268
44	7	1.4	228	22	ABG00053
45	7	1.4	234	16	AAI82605
46	7	1.4	234	24	ABR47768
47	7	1.4	234	24	ABR00092
48	7	1.4	238	16	AAI74981
49	7	1.4	238	23	AAI79439
50	7	1.4	245	23	ABG5468
51	7	1.4	250	22	ABG69076
52	7	1.4	259	21	AAI24855
53	7	1.4	276	18	AAI20822
54	7	1.4	278	22	AAI91401
55	7	1.4	278	22	AAI95498
56	7	1.4	278	22	AAI79157
57	7	1.4	281	22	ABG61929
58	7	1.4	287	22	ABG28799
59	7	1.4	288	22	AAI64161
60	7	1.4	293	23	ABG53671
61	7	1.4	296	21	AAI24854
62	7	1.4	301	21	AAI24853
63	7	1.4	308	22	AAI46546
64	7	1.4	312	22	AAI81921
65	7	1.4	312	22	AAI35466
66	7	1.4	312	22	AAI35471
67	7	1.4	313	21	AAI11138
68	7	1.4	320	23	ABP39594
69	7	1.4	327	20	AAI05529
70	7	1.4	327	21	AAI97832
71	7	1.4	329	23	AAI87041
72	7	1.4	329	23	AAI87042
73	7	1.4	331	21	AAI6831
74	7	1.4	341	23	ABP27788
75	7	1.4	344	23	ABP43788
76	7	1.4	365	21	AAI09073
77	7	1.4	365	21	AAI09073
78	7	1.4	373	23	ABG54724
79	7	1.4	375	23	ABP43792
80	7	1.4	376	22	ABG60164
81	7	1.4	377	22	AAI93690
82	7	1.4	387	23	ABG53661

N. gonorrhoeae ami
Amino acid sequenc
Amino acid sequenc
Bifidobacterium lo
Partial sequence o
Human protein HP10
Human liver peptid
Peptide #7799 enco
Human brain expres
Human bone marrow
Peptide #803 enco
Human peptide enco
Drosophila melanog
Human polypeptide
HIV-1 group O stra
Novel human diagn
Agmenellum quadrip
Sequence homologou
Novel human diagno
Human polypeptide
Lactococcus lactis
Arabidopsis thalia
Eucalyptus grandis
Drosophila melanog
Staphylococcus epi
Wheat branched cha
Human immune/haema
Drosophila melanog
Drosophila melanog
Propionibacterium
Mycobacterium tube
Drosophila melanog
Novel human diagno
Eph transmembrane
Human secreted pro
Human gene 82 enco
Human hex-L protei
Braas/colon cance
Lactococcus lactis
Drosophila melanog
Arabidopsis thalia
H. pylori secreted
C glutaminum prote
Human protein sequ
Corynebacterium gl
Novel human diagno
Novel human diagno
Propionibacterium
Lactococcus lactis
Arabidopsis thalia
Arabidopsis thalia
Propionibacterium
S. epidermidis ope
Escherichia coli E
Bordetella pertuss
Staphylococcus epi
Soybean vestitone
Adenovirus 5 recom
Adenovirus 5 recom
37AD protein #2.
Streptococcus poly
MBP protein. Hom
Arabidopsis thalia
Arabidopsis thalia
Lactococcus lactis
MBP protein #2.
Drosophila melanog
Human polypeptide,
Lactococcus lactis

83	7	1.4	436	17	AAR95636	156	7	1.4	1461	24	ABP57442	Mycobacterium tube
84	7	1.4	438	18	AAW18209	157	7	1.4	1568	23	ABG97467	Human NAAPI, from
85	7	1.4	439	22	AAU00837	158	7	1.4	1594	22	ABB63688	Drosophila melanog
86	7	1.4	440	21	AAZ20592	159	7	1.4	1576	24	ABJ25188	Aspergillus terreu
87	7	1.4	440	21	AAI69545	160	7	1.4	1591	23	AAW52844	Aspergillus fumiga
88	7	1.4	451	21	AAI12576	161	7	1.4	1941	13	ABP65509	Aspergillus medi
89	7	1.4	466	19	AAW84357	162	7	1.4	2705	22	AAAB90772	Blitodabacterium lo
90	7	1.4	472	22	ABG23847	163	7	1.4	3080	14	AAAR35081	Human shear stress
91	7	1.4	478	21	AAI35968	164	7	1.4	3413	19	AAW52849	ZYVW polyprotein..
92	7	1.4	486	23	ABR05347	165	7	1.4	4498	22	ABW58595	A. mediterranei ri
93	7	1.4	489	22	ABW87879	166	7	1.4	7068	22	AAE10142	Drosophila melanog
94	7	1.4	511	22	ABW64688	167	7	1.4	9477	22	AAE10144	Streptomyces nous
95	7	1.4	512	10	AAI94880	168	7	1.4	11096	22	AAE10129	Streptomyces nous
96	7	1.4	512	12	AAI44908	169	6	1.2	6	22	AAW5777	PDZ motif sequence
97	7	1.4	512	12	AAI44909	170	6	1.2	6	22	AAW57609	Neurexin C-termina
98	7	1.4	516	24	ABU01580	171	6	1.2	6	22	AAW58019	Neurexin C-termina
99	7	1.4	519	24	ABP71695	172	6	1.2	6	23	ABJ05275	T-cell surface rec
100	7	1.4	521	9	AAI80930	173	6	1.2	6	23	ABP63503	PDZ motif (PL) pep
101	7	1.4	525	21	AAI09072	174	6	1.2	6	23	AAW64500	Neurotoxic beta-am
102	7	1.4	525	21	AAI09072	175	6	1.2	7	19	AAW64500	PDZ motif sequence
103	7	1.4	525	23	ABR93308	176	6	1.2	7	22	AAW57610	Neurexin C-termina
104	7	1.4	534	20	AAW82789	177	6	1.2	7	22	AAW57610	Neurexin C-termina
105	7	1.4	540	21	AAI43056	178	6	1.2	7	23	ABJ05276	T-cell surface rec
106	7	1.4	544	21	AAI57302	179	6	1.2	7	23	ABP63504	PDZ motif (PL) pep
107	7	1.4	549	21	AAI09071	180	6	1.2	7	23	AAU99458	Dansyl-labeled be
108	7	1.4	549	21	AAI09071	181	6	1.2	8	20	AAW06067	Human cancer anti
109	7	1.4	551	20	AAI73938	182	6	1.2	8	22	AAW55779	PDZ motif sequence
110	7	1.4	552	22	AAI39411	183	6	1.2	8	22	AAW57611	Neurexin C-termina
111	7	1.4	556	22	AAI41197	184	6	1.2	8	22	AAW58021	Neurexin C-termina
112	7	1.4	557	22	AAW14170	185	6	1.2	8	23	ABJ05277	T-cell surface rec
113	7	1.4	557	24	ABU1766	186	6	1.2	8	23	ABP63505	PDZ motif (PL) pep
114	7	1.4	567	18	AAW55703	187	6	1.2	9	15	AAW57742	Protein-kinase inh
115	7	1.4	587	22	AAI94824	188	6	1.2	9	20	AAW06046	Human cancer anti
116	7	1.4	587	23	AAI55554	189	6	1.2	9	20	AAW05968	Human cancer anti
117	7	1.4	611	22	AAI36518	190	6	1.2	9	21	AAI79750	NY-ESO-1 derived p
118	7	1.4	617	18	AAW7247	191	6	1.2	9	24	ABP74471	Human NY-ESO-1 epi
119	7	1.4	659	23	AAU98892	192	6	1.2	10	19	AAW30675	Brassicica napus acy
120	7	1.4	683	22	ABG30138	193	6	1.2	10	20	AAW06069	Human cancer anti
121	7	1.4	687	22	ABR64429	194	6	1.2	10	20	AAW06070	Human cancer anti
122	7	1.4	700	21	AAI75589	195	6	1.2	10	20	AAW06071	Human cancer anti
123	7	1.4	700	23	ABP77608	196	6	1.2	10	20	AAW06072	Human cancer anti
124	7	1.4	747	23	ABW3579	197	6	1.2	10	20	AAW06015	Human cancer anti
125	7	1.4	792	23	AAU83419	198	6	1.2	10	20	AAW05969	Human cancer anti
126	7	1.4	844	19	AAW98721	199	6	1.2	10	20	AAW05970	Human cancer anti
127	7	1.4	859	17	AAW07872	200	6	1.2	10	20	AAW05971	Human cancer anti
128	7	1.4	886	18	AAW93787	201	6	1.2	10	20	AAW05972	Human cancer anti
129	7	1.4	891	22	AAE0586	202	6	1.2	10	20	AAW05973	Human cancer anti
130	7	1.4	895	24	ABJ18910	203	6	1.2	10	20	AAW05974	Human cancer anti
131	7	1.4	909	24	ABP56877	204	6	1.2	10	20	AAW05976	Human cancer anti
132	7	1.4	946	22	AAE00582	205	6	1.2	10	20	AAW05977	Human cancer anti
133	7	1.4	962	21	AAI49944	206	6	1.2	10	22	AAE07734	Human MHC class I
134	7	1.4	988	22	AAE01043	207	6	1.2	10	22	AAI31331	Exemplary antigen
135	7	1.4	988	22	AAI31714	208	6	1.2	10	23	ABG66802	Tumour antigen NY-
136	7	1.4	988	23	ABG70999	209	6	1.2	11	20	AAW06066	Human cancer anti
137	7	1.4	988	23	ABR76165	210	6	1.2	11	20	AAW06068	Human cancer anti
138	7	1.4	988	23	AAI04214	211	6	1.2	11	20	AAW05967	Human cancer anti
139	7	1.4	989	21	AAI49945	212	6	1.2	11	21	AAW99291	HLA class II bindi
140	7	1.4	1018	22	AAE00581	213	6	1.2	11	22	AAE07733	Human MHC class I
141	7	1.4	1047	22	AAW78507	214	6	1.2	11	22	AAE04942	Nuclear Dbp2-relat
142	7	1.4	1047	22	AAW79491	215	6	1.2	11	22	AAE04943	Proteome analysis
143	7	1.4	1095	24	ABJ25381	216	6	1.2	11	24	ABP74798	MHC binding peptid
144	7	1.4	1102	23	ABR93939	217	6	1.2	12	13	AAI21768	MHC binding peptid
145	7	1.4	1174	22	ABW65304	218	6	1.2	12	14	AAI34801	Major histocompati
146	7	1.4	1187	20	AAI05841	219	6	1.2	12	20	AAW06065	Human cancer anti
147	7	1.4	1204	21	AAI48838	220	6	1.2	12	21	AAW99269	HLA class II bindi
148	7	1.4	1216	21	AAI48837	221	6	1.2	12	23	AAU87915	PDZ domain binding
149	7	1.4	1219	22	ABW65388	222	6	1.2	12	23	AAU87916	PDZ domain binding
150	7	1.4	1239	22	ABW61801	223	6	1.2	13	13	AAI21769	MHC binding peptid
151	7	1.4	1248	21	AAI48836	224	6	1.2	13	14	AAI34807	Major histocompati
152	7	1.4	1277	24	ABJ25981	225	6	1.2	13	14	AAI34808	Major histocompati
153	7	1.4	1366	22	ABW66921	226	6	1.2	13	14	AAI34809	Major histocompati
154	7	1.4	1447	21	AAI36514	227	6	1.2	13	14	AAI34810	Major histocompati
155	7	1.4	1447	23	ABP73835	228	6	1.2	13	14	AAI34811	Major histocompati

229	6	1.2	13	14	AAR34812	Major histocompati	302	6	1.2	34	22	ABG59387	Human liver peptid
230	6	1.2	13	14	AAR34813	Major histocompati	303	6	1.2	34	22	ABG23702	Novel human diago
231	6	1.2	13	14	AAR34814	Major histocompati	304	6	1.2	34	22	ABB40088	Peptide #7594 enco
232	6	1.2	13	14	AAR34815	Major histocompati	305	6	1.2	34	22	ABB44000	Peptide #11506 enc
233	6	1.2	13	14	AAR34816	Major histocompati	306	6	1.2	34	22	ABB24572	Protein #6571 enco
234	6	1.2	13	14	AAR34817	Major histocompati	307	6	1.2	34	22	ABB25908	Protein #8907 enco
235	6	1.2	13	14	AAR34818	Major histocompati	308	6	1.2	34	22	AAH60845	Human brain expres
236	6	1.2	13	14	AAR34819	Major histocompati	309	6	1.2	34	22	AAH65014	Human brain expres
237	6	1.2	13	14	AAR34834	Major histocompati	310	6	1.2	34	22	AAH73528	Human bone marrow
238	6	1.2	13	14	AAR34836	Major histocompati	311	6	1.2	34	22	AAH77731	Human bone marrow
239	6	1.2	13	14	AAR34837	Major histocompati	312	6	1.2	34	22	AAH20001	Peptide #6435 enco
240	6	1.2	13	14	AAR34838	Major histocompati	313	6	1.2	34	22	AAH21640	Peptide #8074 enco
241	6	1.2	13	14	AAR34839	Major histocompati	314	6	1.2	34	22	AAH33722	Peptide #7759 enco
242	6	1.2	13	14	AAR34820	Major histocompati	315	6	1.2	34	22	AAH37945	Peptide #11982 enc
243	6	1.2	13	14	AAR34821	Major histocompati	316	6	1.2	34	22	AAH99761	ERA binding domain
244	6	1.2	13	14	AAR34822	Major histocompati	317	6	1.2	34	23	ABG43389	Human peptid enco
245	6	1.2	13	14	AAR34823	Major histocompati	318	6	1.2	34	22	ABG46761	Human peptid enco
246	6	1.2	13	14	AAR34824	Major histocompati	319	6	1.2	37	22	ABB02689	Vital core polypep
247	6	1.2	13	14	AAR34825	Major histocompati	320	6	1.2	38	22	AAH99869	ERA binding domain
248	6	1.2	13	14	AAR34826	Major histocompati	321	6	1.2	38	23	ABB82354	Pufferfish urococt
249	6	1.2	13	14	AAR34827	Major histocompati	322	6	1.2	38	23	ABP25992	Streptococcus poly
250	6	1.2	13	14	AAR34828	Major histocompati	323	6	1.2	39	21	AAH98830	Core polypeptide f
251	6	1.2	13	14	AAR34829	Major histocompati	324	6	1.2	39	22	ABB01238	Vital Dp18/107-1i
252	6	1.2	13	14	AAR34830	Major histocompati	325	6	1.2	39	22	AAH13784	Dp178-1like/Dp107-1
253	6	1.2	13	14	AAR34831	Major histocompati	326	6	1.2	39	22	AAH78231	Core polypeptide T
254	6	1.2	13	14	AAR34832	Major histocompati	327	6	1.2	39	23	ABP53558	Plasmid PRSET Sfil
255	6	1.2	13	14	AAR34833	Major histocompati	328	6	1.2	39	23	ABP53560	Plasmid PRSET Sfil
256	6	1.2	13	18	AAW22116	Padre (pan-DR bind	329	6	1.2	39	23	ABP53562	Plasmid PRSET Sfil
257	6	1.2	13	19	AAW50121	Pan DR binding pep	330	6	1.2	40	22	AAH58234	Human brain expres
258	6	1.2	13	20	AAH06064	Human cancer antiq	331	6	1.2	40	22	AAH99833	ERA binding domain
259	6	1.2	13	21	AAH26807	Phosphoryl group a	332	6	1.2	42	22	AAH99788	ERA binding domain
260	6	1.2	13	21	AAH99270	HLA class II bindi	333	6	1.2	42	22	AAH99788	ERA binding domain
261	6	1.2	13	23	ABP54999	Cyclin dependent p	334	6	1.2	43	20	AAH82294	Dynamn 41.2 ZELan
262	6	1.2	13	24	ABH97973	Peptide used as ph	335	6	1.2	43	23	AAH015500	Porphyrin yezoensis
263	6	1.2	13	24	ABH56628	Cyclin dependent p	336	6	1.2	44	22	AAH08629	Human polypeptide
264	6	1.2	13	24	ABH59484	Phosphoryl group a	337	6	1.2	46	22	ABG01742	Novel human diago
265	6	1.2	13	24	ABH55027	Cyclin dependent p	338	6	1.2	46	22	AAH99806	ERA binding domain
266	6	1.2	14	14	AAR34835	Major histocompati	339	6	1.2	46	22	AAH99854	ERA binding domain
267	6	1.2	14	18	AAH15095	Anti-HIV p24 antib	340	6	1.2	47	21	AAH09429	Hepatitis GB virus
268	6	1.2	14	20	AAH05986	Human cancer antiq	341	6	1.2	47	22	AAH99791	ERA binding domain
269	6	1.2	14	22	ABP25107	Human MHC peptide	342	6	1.2	47	22	AAH99800	ERA binding domain
270	6	1.2	14	22	AAH00229	Human eph fragment	343	6	1.2	48	20	AAH60201	Human endometrium
271	6	1.2	14	22	AAH00231	Human eph fragment	344	6	1.2	49	22	ABG50143	Human liver peptid
272	6	1.2	14	22	AAH03837	Hepatitis C virus	345	6	1.2	49	22	ABB30100	Peptide #2751 enco
273	6	1.2	14	23	ABH09705	Hepatitis B virus	346	6	1.2	49	22	ABB35271	Peptide #2777 enco
274	6	1.2	15	20	AAH05978	Human cancer antiq	347	6	1.2	49	22	ABB20713	Protein #2712 enco
275	6	1.2	19	22	AAH55819	PL peptide #30. S	348	6	1.2	49	22	AAH56102	Human brain expres
276	6	1.2	19	22	AAH57681	Neurexin PL peptid	349	6	1.2	49	22	AAH68472	Human bone marrow
277	6	1.2	19	22	AAH58091	Neurexin PL peptid	350	6	1.2	49	22	AAH16281	Peptide #2715 enco
278	6	1.2	19	23	ABH05211	C-terminal peptide	351	6	1.2	49	22	AAH28768	Peptide #2805 enco
279	6	1.2	19	23	ABH61335	Neurexin PL peptid	352	6	1.2	49	22	AAH99770	ERA binding domain
280	6	1.2	20	20	AAH05979	Human cancer antiq	353	6	1.2	49	22	AAH99899	ERA binding domain
281	6	1.2	20	20	AAH82295	Dynamn 41.20 ZELa	354	6	1.2	49	22	AAH04014	Peptide #2696 enco
282	6	1.2	22	24	AAH01623	Agrobacterium tune	355	6	1.2	49	23	ABG38053	Human peptid enco
283	6	1.2	23	20	AAH82310	Transport peptide	356	6	1.2	50	22	AAH42299	Propionibacterium
284	6	1.2	23	22	AAH98276	Anti-A33 antigen i	357	6	1.2	50	22	AAH99842	ERA binding domain
285	6	1.2	23	22	AAH98277	Anti-A33 antigen i	358	6	1.2	51	22	AAH06820	Human foetal prote
286	6	1.2	23	22	AAH98278	Anti-A33 antigen i	359	6	1.2	51	22	AAH06825	Human foetal prote
287	6	1.2	23	22	AAH97656	A33 antigen bindin	360	6	1.2	52	21	AAH64748	Human 5' EST relat
288	6	1.2	23	22	AAH97657	A33 antigen bindin	361	6	1.2	52	22	AAH99764	ERA binding domain
289	6	1.2	23	22	AAH97658	A33 antigen bindin	362	6	1.2	53	19	AAH38647	S. pneumoniae ATP
290	6	1.2	25	22	ABG54356	Human liver peptid	363	6	1.2	53	22	AAH42426	Propionibacterium
291	6	1.2	25	22	ABH19377	Peptide #6883 enco	364	6	1.2	54	22	AAH56774	Propionibacterium
292	6	1.2	25	22	ABB24164	Protein #6163 enco	365	6	1.2	54	22	AAH99848	ERA binding domain
293	6	1.2	25	22	AAH60055	Human brain expres	366	6	1.2	54	22	AAH77023	Mutant GHR exon 7
294	6	1.2	25	22	AAH72659	Human bone marrow	367	6	1.2	55	16	AAH77020	Mutant GHR exon 7
295	6	1.2	25	22	AAH19670	Peptide #6104 enco	368	6	1.2	55	16	AAH77021	Mutant GHR exon 7
296	6	1.2	25	22	AAH32892	Peptide #6929 enco	369	6	1.2	55	16	AAH77022	Human GHR exon 7 p
297	6	1.2	25	23	ABG62600	Bubacterial DnaH1	370	6	1.2	55	18	AAH44242	Growth hormone rec
298	6	1.2	25	23	ABG42483	Human peptid enco	371	6	1.2	55	18	AAH44243	Growth hormone rec
299	6	1.2	30	23	AAH85105	Human NIN501a segm	372	6	1.2	55	18	AAH44244	Growth hormone rec
300	6	1.2	31	20	AAH70858	Human Zcyto5 epit	373	6	1.2	55	18	AAH44245	Growth hormone rec
301	6	1.2	34	22	ABG55255	Human liver peptid	374	6	1.2	55	19	AAH70558	Growth hormone rec

375	6	1.2	55	19	AAW70559	Growth hormone rec	448	6	1.2	70	22	AAW19372	Peptide #5806 enco
376	6	1.2	55	19	AAW70560	Growth hormone rec	449	6	1.2	70	22	AAW32224	Peptide #6261 enco
377	6	1.2	55	19	AAW70561	Growth hormone rec	450	6	1.2	70	22	ABG41760	Human peptide enco
378	6	1.2	55	22	ABG08147	Novel human diagno	451	6	1.2	71	22	AAU44830	Proionibacterium
379	6	1.2	55	22	ABR17780	Human nervous syst	452	6	1.2	71	22	AAW9653	ERA binding domain
380	6	1.2	55	23	ABP05003	Human ORFX protein	453	6	1.2	71	22	AAW96534	ERA binding domain
381	6	1.2	56	22	ABG29048	Novel human diagno	454	6	1.2	71	22	AAU99186	Target molecule hu
382	6	1.2	57	22	ABG91744	C glutamicum prote	455	6	1.2	72	22	ABG59207	Human liver peptid
383	6	1.2	57	23	ABP32584	Human ORF1557 prot	456	6	1.2	72	22	AAU49195	Proionibacterium
384	6	1.2	58	22	ABR6408	Human testicular a	457	6	1.2	72	22	ABG08844	Novel human diagno
385	6	1.2	58	22	AAU47762	Proionibacterium	458	6	1.2	72	22	AAW64808	Human brain expres
386	6	1.2	58	22	AAW5877	Human reproductiv	459	6	1.2	72	22	AAW77564	Human bone marrow
387	6	1.2	58	22	AAW92725	Human digestive sy	460	6	1.2	72	22	AAW02535	Human polypeptide
388	6	1.2	58	22	AAW99860	ERA binding domain	461	6	1.2	72	22	AAW37746	Peptide #11783 enc
389	6	1.2	58	22	AAW99866	ERA binding domain	462	6	1.2	72	22	ABG46592	Human peptide enco
390	6	1.2	59	21	AAW33846	Arabidopsis thalia	463	6	1.2	73	21	AAW52618	Helicobacter pylor
391	6	1.2	59	22	ABG52518	Human liver peptid	464	6	1.2	73	22	ABW96455	Novel testicular a
392	6	1.2	59	22	ABG58236	Human liver peptid	465	6	1.2	73	22	ABG25287	Novel human diagno
393	6	1.2	59	22	AAW56677	Proionibacterium	466	6	1.2	73	22	AAW95924	Human reproductiv
394	6	1.2	59	22	ABR42830	Peptide #10336 enc	467	6	1.2	73	22	AAW51054	Helicobacter pylor
395	6	1.2	59	22	AAW62611	Protein #8110 enco	468	6	1.2	74	13	AAW30165	UGT1 Exon 4 produc
396	6	1.2	59	22	AAW63722	Human brain expres	469	6	1.2	74	21	AAW12471	Zea mays protein f
397	6	1.2	59	22	AAW76536	Human bone marrow	470	6	1.2	74	22	AAW54587	Proionibacterium
398	6	1.2	59	22	AAW20926	Peptide #7360 enco	471	6	1.2	75	16	AAW71130	SH2 domain of cyto
399	6	1.2	59	22	AAW36643	Peptide #10680 enc	472	6	1.2	75	21	AAW34737	Human secreted pro
400	6	1.2	59	22	AAW99662	ERA binding domain	473	6	1.2	75	21	AAW01935	Human secreted pro
401	6	1.2	59	22	AAW99704	ERA binding domain	474	6	1.2	75	21	AAW87587	Aeromonas caviae a
402	6	1.2	59	22	AAW99719	ERA binding domain	475	6	1.2	75	22	ABG09773	Novel human diagno
403	6	1.2	59	22	AAW99725	ERA binding domain	476	6	1.2	75	23	ABP04130	Human ORFX protein
404	6	1.2	59	22	AAW99812	ERA binding domain	477	6	1.2	76	23	ABP08002	Human ORFX protein
405	6	1.2	59	22	AAW99818	ERA binding domain	478	6	1.2	77	16	AAW82858	Human ORFX protein
406	6	1.2	59	22	AAW99890	ERA binding domain	479	6	1.2	77	23	ABW81200	Fragment of alpha-
407	6	1.2	59	22	ABG45789	Human peptide enco	480	6	1.2	78	22	AAW55449	E. coli Z-ring int
408	6	1.2	60	22	AAW99661	ERA binding domain	481	6	1.2	78	17	AAW90781	Proionibacterium
409	6	1.2	60	22	AAW99663	ERA binding domain	482	6	1.2	79	23	ABP31043	Reverse-FRAME anti
410	6	1.2	60	22	AAW99713	ERA binding domain	483	6	1.2	79	23	ABP05175	Human ORF16 protei
411	6	1.2	60	22	AAW99731	ERA binding domain	484	6	1.2	80	17	AAW90788	Human ORFX protein
412	6	1.2	60	22	AAW99773	ERA binding domain	485	6	1.2	80	17	AAW90787	Reverse-FRAME anti
413	6	1.2	60	22	AAW99827	ERA binding domain	486	6	1.2	80	17	AAW90786	Reverse-FRAME anti
414	6	1.2	60	22	AAW99845	ERA binding domain	487	6	1.2	80	17	AAW90785	Reverse-FRAME anti
415	6	1.2	60	23	ABP42201	Human ovarian anti	488	6	1.2	80	17	AAW90784	Reverse-FRAME anti
416	6	1.2	60	24	ABP97608	Amino acid sequenc	489	6	1.2	80	17	AAW90783	Reverse-FRAME anti
417	6	1.2	62	22	ABG19857	Novel human diagno	490	6	1.2	80	17	AAW90782	Reverse-FRAME anti
418	6	1.2	63	22	ABG20392	Novel human diagno	491	6	1.2	80	22	ABG50283	Human liver peptid
419	6	1.2	63	22	ABG26590	Novel haematologic	492	6	1.2	80	22	AAW66913	Proionibacterium
420	6	1.2	64	22	AAW81428	Human haematologic	493	6	1.2	80	22	ABW30255	Peptide #2906 enco
421	6	1.2	64	22	AAW81692	Human haematologic	494	6	1.2	80	22	ABW35421	Peptide #2927 enco
422	6	1.2	64	23	ABP34873	Human kinase-like	495	6	1.2	80	22	ABW20858	Peptide #2857 enco
423	6	1.2	65	20	AAW23796	A gyrase protein s	496	6	1.2	80	22	AAW78220	Human mitochondria
424	6	1.2	65	22	AAU41751	Proionibacterium	497	6	1.2	80	22	AAW56245	Human brain expres
425	6	1.2	65	22	AAU51508	Proionibacterium	498	6	1.2	80	22	AAW68623	Human bone marrow
426	6	1.2	65	22	AAW56563	Human brain expres	499	6	1.2	80	22	AAW16430	Peptide #2864 enco
427	6	1.2	65	22	AAW78269	Human bone marrow	500	6	1.2	80	22	AAW28929	Peptide #2966 enco
428	6	1.2	65	22	AAW05339	Human polypeptide	501	6	1.2	80	22	AAW04164	Peptide #2846 enco
429	6	1.2	65	23	ABG47295	Human peptide enco	502	6	1.2	80	23	ABG38201	Human peptide enco
430	6	1.2	65	23	ABP07279	Human ORFX protein	503	6	1.2	81	17	AAW90789	Reverse-FRAME anti
431	6	1.2	66	21	AAW83801	Human secreted pro	504	6	1.2	81	24	AAW79708	Portion of NM 0033
432	6	1.2	66	22	ABG20253	Novel human diagno	505	6	1.2	81	24	AAW79709	Portion of UB2467
433	6	1.2	66	22	ABW4888	Human secreted pro	506	6	1.2	81	24	AAW79711	Portion of mouse 5
434	6	1.2	66	23	AAO15501	Porphyra yezoensis	507	6	1.2	81	24	AAW79712	Portion of US2433
435	6	1.2	68	21	AAW12472	Zea mays protein f	508	6	1.2	81	24	AAW79713	Portion of US2824
436	6	1.2	68	22	AAU43126	Proionibacterium	509	6	1.2	82	22	AAW56643	Proionibacterium
437	6	1.2	69	21	AAW40541	Human ORFX ORF305	510	6	1.2	82	22	AAW64588	Proionibacterium
438	6	1.2	69	21	AAW37128	Arabidopsis thalia	511	6	1.2	82	22	AAW66340	Proionibacterium
439	6	1.2	69	22	ABW3698	Human gastric canc	512	6	1.2	83	22	ABW41733	Peptide #2939 enco
440	6	1.2	69	23	ABP04101	Human ORFX protein	513	6	1.2	83	22	AAW62606	Human brain expres
441	6	1.2	69	23	ABP05881	Human ORFX protein	514	6	1.2	83	22	AAW03013	Human polypeptide
442	6	1.2	70	22	ABG53629	Human liver peptid	515	6	1.2	83	22	AAW35528	Peptide #9565 enco
443	6	1.2	70	22	ABW838751	Peptide #6257 enco	516	6	1.2	83	23	ABP01772	Human ORFX protein
444	6	1.2	70	22	ABW23797	Protein #5796 enco	517	6	1.2	84	22	ABG02717	Novel human diagno
445	6	1.2	70	22	AAW59390	Human brain expres	518	6	1.2	85	22	AAW69450	Human purified sec
446	6	1.2	70	22	AAW71945	Human bone marrow	519	6	1.2	85	22	AAW65806	RP55-like protein
447	6	1.2	70	22	AAW08876	Human polypeptide	520	6	1.2	85	23	AAW15498	Porphyra yezoensis

521	6	1.2	86	22	AAM91998	Human digestive sy	594	6	1.2	109	21	AAG34350	Zea mays protein f
522	6	1.2	87	23	ABP64976	Human protein SEQ	595	6	1.2	109	22	AAO09868	Human polypeptide
523	6	1.2	87	22	ABG99699	ERA binding domain	596	6	1.2	109	22	AAO98230	Rabbit anti-A33 an
524	6	1.2	88	22	ABG23994	Novel human diapo	597	6	1.2	109	22	AAO98231	Rabbit anti-A33 an
525	6	1.2	88	22	AAO08786	Human polypeptide	598	6	1.2	109	22	AAO98232	Rabbit anti-A33 an
526	6	1.2	89	24	AAO28648	Maize basel layer	599	6	1.2	109	22	AAO75149	Rabbit anti-A33 an
527	6	1.2	89	24	AAO28649	Maize basel layer	600	6	1.2	109	22	AAO75150	Rabbit anti A33 an
528	6	1.2	90	22	AAO28649	Maize basel layer	601	6	1.2	109	22	AAO75150	Rabbit anti A33 an
529	6	1.2	90	24	AAO28647	Maize BTL-2 prote	602	6	1.2	110	15	AAO56390	Extracellular hGR
530	6	1.2	90	24	AAO28647	Maize protein, SEQ	603	6	1.2	110	23	AAO15499	Porphyra yezoensis
531	6	1.2	91	21	AAO18124	Arabidopsis thalia	604	6	1.2	111	21	AAO54944	Arabidopsis thalia
532	6	1.2	91	21	AAO32699	Zea mays protein f	605	6	1.2	112	15	AAO56391	Extracellular hGR
533	6	1.2	91	22	AAO56732	Proteinibacterium	606	6	1.2	112	22	AAO60402	Drosophila melanog
534	6	1.2	92	23	ABP00953	Human ORFX protein	607	6	1.2	113	22	AAO01362	Human polypeptide
535	6	1.2	92	21	AAO53721	Human ORFX cancer	608	6	1.2	113	23	ABP05606	Human ORFX protein
536	6	1.2	94	22	AAU46786	Proteinibacterium	609	6	1.2	115	21	AAO41132	Zea mays protein f
537	6	1.2	94	22	ABG19735	Novel human diapo	610	6	1.2	115	22	AAO00563	Human polypeptide
538	6	1.2	94	23	ABP28987	Streptococcus poly	611	6	1.2	115	22	AAO02810	Human polypeptide
539	6	1.2	95	21	AAO32844	Eucalyptus grandis	612	6	1.2	116	21	AAO33958	Arabidopsis thalia
540	6	1.2	95	21	AAO61506	Arabidopsis thalia	613	6	1.2	116	21	AAO60235	Novel human diapo
541	6	1.2	95	21	AAO61535	Arabidopsis thalia	614	6	1.2	116	22	ABG02936	Novel human diapo
542	6	1.2	95	22	ABG21313	Novel human diapo	615	6	1.2	117	16	AAO82569	Wheat glycogenin c
543	6	1.2	95	22	ABG16615	Human nervous syst	616	6	1.2	117	22	AAU35520	Haemophilus influe
544	6	1.2	95	23	ABP47811	SMRV protein relat	617	6	1.2	117	22	AAO52520	Wheat water stress
545	6	1.2	96	22	AAU58034	Proteinibacterium	618	6	1.2	117	23	AAU76204	Wheat water stress
546	6	1.2	96	22	ABG21324	Novel human diapo	619	6	1.2	118	22	ABG20410	Novel human diapo
547	6	1.2	97	23	ABP47808	MPV protein relat	620	6	1.2	118	22	AAO02937	Human polypeptide
548	6	1.2	97	23	ABP47809	SRV-2 protein rela	621	6	1.2	119	22	AAO06904	Human polypeptide
549	6	1.2	97	23	ABP47810	OPAV protein relat	622	6	1.2	120	22	ABG89277	Drosophila melanog
550	6	1.2	97	23	ABP47812	MTV protein relat	623	6	1.2	121	21	AAO01085	Human secreted pro
551	6	1.2	97	23	ABP47813	Listeria monocytog	624	6	1.2	121	21	AAO01086	Human secreted pro
552	6	1.2	97	23	ABG49706	Novel human diapo	625	6	1.2	121	21	AAO01087	Human secreted pro
553	6	1.2	98	21	AAO61534	Arabidopsis thalia	626	6	1.2	121	21	AAO01088	Human secreted pro
554	6	1.2	98	21	AAO01969	Human secreted pro	627	6	1.2	121	21	AAO01089	Human secreted pro
555	6	1.2	98	22	AAU48448	Proteinibacterium	628	6	1.2	121	22	ABG06733	Novel human diapo
556	6	1.2	99	21	AAO41554	Arabidopsis thalia	629	6	1.2	122	22	ABG29522	Novel human diapo
557	6	1.2	99	22	ABG22584	Novel human diapo	630	6	1.2	122	23	ABP01617	Human ORFX protein
558	6	1.2	99	22	ABG25273	Novel human diapo	631	6	1.2	123	21	AAU93964	Mouse TANGO 219 pr
559	6	1.2	99	22	AAU33019	Novel human secret	632	6	1.2	123	23	AAE26441	Arabidopsis thalia
560	6	1.2	100	20	AAU29205	Amino acid sequenc	633	6	1.2	124	21	AAO24036	Human secreted pro
561	6	1.2	100	22	AAU60043	Proteinibacterium	634	6	1.2	124	23	ABU10908	Human secreted pro
562	6	1.2	100	22	AAU66032	Proteinibacterium	635	6	1.2	126	14	AAO50054	Human secreted pro
563	6	1.2	101	21	AAO09245	Arabidopsis thalia	636	6	1.2	127	21	AAO60234	Arabidopsis thalia
564	6	1.2	101	22	AAU59816	Proteinibacterium	637	6	1.2	127	22	AAU51501	Proteinibacterium
565	6	1.2	101	22	AAU79316	Human ORFX ORF329	638	6	1.2	127	22	AAU93585	Human polypeptide,
566	6	1.2	101	23	ABP64398	Human ORF768. Hom	639	6	1.2	127	23	ABG96450	Human ovarian canc
567	6	1.2	101	23	ABG67068	Streptavidin-BindI	640	6	1.2	128	21	AAO39172	Human secreted pro
568	6	1.2	101	23	ABP05739	Human ORFX protein	641	6	1.2	128	22	AAO99659	ERA binding domain
569	6	1.2	101	24	ABU57590	Streptavidin-BindI	642	6	1.2	128	23	ABP69240	Human polypeptide
570	6	1.2	102	21	AAO40565	Human ORFX ORF329	643	6	1.2	129	21	AAO10137	Arabidopsis thalia
571	6	1.2	102	21	AAO10138	Arabidopsis thalia	644	6	1.2	129	21	AAO61177	Arabidopsis thalia
572	6	1.2	102	21	AAO57676	Arabidopsis thalia	645	6	1.2	129	22	AAU33662	Pseudomonas aerugi
573	6	1.2	102	21	AAO61178	Arabidopsis thalia	646	6	1.2	129	22	AAU42678	Proteinibacterium
574	6	1.2	102	22	AAO41997	Human polypeptide	647	6	1.2	129	22	AAO93833	Human protein sequ
575	6	1.2	102	23	ABP32691	Human ORF1664 prot	648	6	1.2	129	24	ABP81256	Arabidopsis thalia
576	6	1.2	103	24	ABP58239	Human ovary-specif	649	6	1.2	130	22	AAU33503	Enterococcus faeca
577	6	1.2	105	23	ABP64328	Human ORF698. Hom	650	6	1.2	130	22	AAU37768	Streptococcus pneu
578	6	1.2	106	17	AAO90791	Reverse-Frame anti	651	6	1.2	130	22	AAU38087	Streptococcus pneu
579	6	1.2	106	21	AAO40735	Human ORFX ORF499	652	6	1.2	130	23	ABP25699	Streptococcus poly
580	6	1.2	106	23	ABP02646	Human ORFX protein	653	6	1.2	130	24	ABP81545	Streptococcus pneu
581	6	1.2	107	21	AAO60236	Arabidopsis thalia	654	6	1.2	130	24	ABU00644	S. pneumoniae type
582	6	1.2	107	22	AAO83817	Human immune/haema	655	6	1.2	131	21	AAO51671	Gene 1 human secre
583	6	1.2	107	23	AAO18390	Murine CBRI1. light	656	6	1.2	131	21	AAO51672	Human secreted pro
584	6	1.2	107	23	AAO18392	Humanised murine C	657	6	1.2	131	22	AAU35463	Haemophilus influe
585	6	1.2	107	23	AAO18393	Humanised murine C	658	6	1.2	131	22	AAO99997	Human protein sequ
586	6	1.2	107	23	AAO18394	Humanised murine C	659	6	1.2	131	23	ABG93340	C. albicans BAX-as
587	6	1.2	107	24	ABP76286	Human GENSER prote	660	6	1.2	132	22	AAU87477	Novel central nerv
588	6	1.2	108	22	AAU34776	E. coli cellular p	661	6	1.2	132	22	AAU54733	Proteinibacterium
589	6	1.2	108	22	AAU38211	Salmonella typhi c	662	6	1.2	132	22	AAU19850	Human novel extrac
590	6	1.2	108	22	ABP17577	Human nervous syst	663	6	1.2	132	23	ABP48070	Human polypeptide
591	6	1.2	108	23	ABP29635	Streptococcus poly	664	6	1.2	133	22	ABG09248	Novel human diapo
592	6	1.2	109	19	AAO38664	S. pneumoniae 30S	665	6	1.2	133	23	AAE25926	Corn GDP synthase
593	6	1.2	109	21	AAO33959	Arabidopsis thalia	666	6	1.2	133	24	ABU67323	Corn geranylgerany

667	6	1.2	134	9	AAp81139	Sequence of plant	740	6	1.2	156	22	ABG15620	Novel human diagno
668	6	1.2	134	13	AAr28465	Sequence encoded b	741	6	1.2	157	19	AAW56252	Interleukin-13 bin
669	6	1.2	134	15	AAr54979	Brassic seed acyl	742	6	1.2	157	22	AAW86952	Human DNA repair a
670	6	1.2	134	16	AAr76695	B. campestris ACP	743	6	1.2	157	22	ABG01094	Novel human diagno
671	6	1.2	134	18	AAW15416	ACP protein encode	744	6	1.2	157	22	ABBS3058	Escherichia coli p
672	6	1.2	134	19	AAW30527	Brassica campestr	745	6	1.2	157	22	AAU21724	Novel human neopla
673	6	1.2	134	19	AAW30671	Brassica campestr	746	6	1.2	157	22	AAU21834	Novel human neopla
674	6	1.2	134	21	AAW49927	Partial genomic AC	747	6	1.2	158	22	AAU87210	Novel central nerv
675	6	1.2	134	22	AAW85932	Amino acid sequenc	748	6	1.2	158	22	AAU87520	Novel central nerv
676	6	1.2	134	23	ABH89266	Human polypeptide	749	6	1.2	158	22	AAU17724	Novel human respir
677	6	1.2	135	21	AAW57675	Arabidopsis thalia	750	6	1.2	158	22	AAU19727	Human novel extrac
678	6	1.2	136	21	AAW5074	Arabidopsis thalia	751	6	1.2	158	22	ABP47947	Human polypeptide
679	6	1.2	136	22	AAW58616	Propionibacterium	752	6	1.2	159	21	AAW4784	Human ORFX ORF1548
680	6	1.2	136	22	ABG28035	Novel human diagno	753	6	1.2	159	21	AAW71041	Streptococcus pyog
681	6	1.2	136	22	AAW84105	Human immune/haema	754	6	1.2	161	21	ABW65804	RPS5-like protein
682	6	1.2	136	22	AAW42362	Human polypeptide	755	6	1.2	161	23	ABP32415	Human transcriptio
683	6	1.2	137	21	AAW16637	Arabidopsis thalia	756	6	1.2	162	23	ABW05775	M. tuberculosis an
684	6	1.2	137	21	AAW43076	Arabidopsis thalia	757	6	1.2	162	23	ABW53404	Lactococcus lactis
685	6	1.2	137	21	ABW11435	Human CIRP-1 homol	758	6	1.2	163	20	AAW41233	A. chroococcum YH2
686	6	1.2	137	24	ABP79973	N. gonorrhoeae amf	759	6	1.2	163	21	AAW40557	Human ORFX ORF321
687	6	1.2	138	18	AAW55485	H. pylori ORF hpj	760	6	1.2	163	21	AAW43062	Human ORFX ORF2826
688	6	1.2	138	22	AAW54545	Propionibacterium	761	6	1.2	163	22	AAW89040	Human immune/haema
689	6	1.2	138	22	AAW90088	C glutamicum proce	762	6	1.2	164	21	AAW7867	M. tuberculosis an
690	6	1.2	138	22	AAW01212	Novel human membra	763	6	1.2	164	22	ABG14120	Novel human diagno
691	6	1.2	138	22	AAW95866	Human protein sequ	764	6	1.2	164	22	AAW71392	Human gene 23-enco
692	6	1.2	138	23	ABP26151	Streptococcus poly	765	6	1.2	164	23	ABW51280	Helicobacter pylor
693	6	1.2	140	21	AAW44225	Human cancer assoc	766	6	1.2	164	24	ABW80985	N. gonorrhoeae amf
694	6	1.2	140	21	AAW57674	Arabidopsis thalia	767	6	1.2	165	21	AAW52613	Helicobacter pylor
695	6	1.2	140	21	AAW01215	Human secreted pro	768	6	1.2	165	21	AAW09965	Arabidopsis thalia
696	6	1.2	140	22	AAW82572	Human immune/haema	769	6	1.2	165	21	AAW50453	Arabidopsis thalia
697	6	1.2	141	21	AAW71040	Streptococcus pyog	770	6	1.2	165	23	ABW92888	Herbicideally activ
698	6	1.2	141	22	AAW52928	Human transmembran	771	6	1.2	166	19	AAW59153	L. lactis rhnb C-t
699	6	1.2	142	22	AAW56689	Propionibacterium	772	6	1.2	166	19	AAW59162	L. lactis MG316 r
700	6	1.2	142	23	AAW25927	Corn GGPP synthase	773	6	1.2	166	21	AAW10731	Arabidopsis thalia
701	6	1.2	142	23	ABW90131	Human polypeptide	774	6	1.2	166	22	AAW63770	Human prostate can
702	6	1.2	142	24	ABW67324	Corn geranylgerany	775	6	1.2	166	22	AAW63812	Human prostate can
703	6	1.2	143	21	AAW41131	Zea mays protein f	776	6	1.2	166	23	ABP29017	Streptococcus poly
704	6	1.2	143	22	ABG19733	Novel human diagno	777	6	1.2	167	21	AAW09244	Arabidopsis thalia
705	6	1.2	143	22	AAW02945	Human polypeptide	778	6	1.2	167	21	AAW15940	E. coli proliferat
706	6	1.2	143	23	ABP01185	Human ORFX protein	779	6	1.2	167	21	AAW71046	Streptococcus pyog
707	6	1.2	144	21	AAW09966	Arabidopsis thalia	780	6	1.2	167	21	AAW41298	Neisseria chimeric
708	6	1.2	144	21	AAW50454	Arabidopsis thalia	781	6	1.2	167	22	AAW85153	Novel cytochrome p
709	6	1.2	144	22	ABG18607	Novel human diagno	782	6	1.2	167	23	ABW52144	Zebrafish P450RA1-
710	6	1.2	144	22	AAW16604	Human olfactory re	783	6	1.2	167	23	ABW94077	Human secreted pro
711	6	1.2	146	21	AAW24791	Plant SDF encoded	784	6	1.2	168	22	ABG20391	Novel human diagno
712	6	1.2	146	21	AAW24954	Human novel secret	785	6	1.2	170	21	AAW07548	Arabidopsis thalia
713	6	1.2	146	22	AAW16613	Human novel secret	786	6	1.2	170	21	AAW09576	Arabidopsis thalia
714	6	1.2	146	24	ABW55682	Human novel polype	787	6	1.2	170	21	AAW52219	Arabidopsis thalia
715	6	1.2	148	19	AAW98432	H. pylori GHP 447	788	6	1.2	170	21	AAW57728	Arabidopsis thalia
716	6	1.2	148	22	AAW06229	Human polypeptide	789	6	1.2	170	22	ABG20255	Novel human diagno
717	6	1.2	149	22	AAW64547	Human ribosome pro	790	6	1.2	171	22	ABG00650	Novel human diagno
718	6	1.2	149	22	AAW14202	Human novel protei	791	6	1.2	171	22	ABG16193	Novel human diagno
719	6	1.2	150	22	AAW86200	Purative P. abyssi	792	6	1.2	171	22	ABG23987	Novel human diagno
720	6	1.2	151	21	AAW03418	Rice putative carb	793	6	1.2	172	21	AAW41678	Human ORFX ORF142
721	6	1.2	151	22	AAW32409	Novel human secret	794	6	1.2	173	21	AAW25395	Pinus radiata cell
722	6	1.2	151	22	AAW07440	Human polypeptide	795	6	1.2	173	22	ABG27583	Novel human diagno
723	6	1.2	151	24	ABR40711	Oryza sativa oil t	796	6	1.2	173	22	AAW38749	Novel subtilisin h
724	6	1.2	152	22	AAW79777	Corynebacterium gl	797	6	1.2	173	22	AAW38776	Novel subtilisin h
725	6	1.2	152	22	AAW79945	Corynebacterium gl	798	6	1.2	173	22	AAW38780	Novel subtilisin h
726	6	1.2	152	23	ABW56528	Bm heparan sulfate	799	6	1.2	173	22	AAW38783	Novel subtilisin h
727	6	1.2	152	24	ABP81307	Streptococcus pneu	800	6	1.2	173	22	AAW38799	Novel subtilisin h
728	6	1.2	153	21	AAW10136	Arabidopsis thalia	801	6	1.2	173	22	AAW38802	Novel subtilisin h
729	6	1.2	153	22	ABG20413	Novel human diagno	802	6	1.2	173	22	AAW38807	Novel subtilisin h
730	6	1.2	153	22	ABW12363	Human bone marrow	803	6	1.2	173	22	AAW38808	Novel subtilisin h
731	6	1.2	153	22	AAW30604	Novel human secret	804	6	1.2	173	22	AAW38824	Novel subtilisin h
732	6	1.2	153	22	AAW95559	ERA binding domain	805	6	1.2	173	22	AAW38827	Novel subtilisin h
733	6	1.2	154	23	ABW92249	Herbicideally activ	806	6	1.2	173	22	AAW38845	Novel subtilisin h
734	6	1.2	155	21	AAW28426	Arabidopsis thalia	807	6	1.2	173	22	AAW38846	Novel subtilisin h
735	6	1.2	155	21	AAW42461	Arabidopsis thalia	808	6	1.2	173	22	AAW38848	Novel subtilisin h
736	6	1.2	155	21	AAW71044	Streptococcus pyog	809	6	1.2	173	22	AAW38854	Novel subtilisin h
737	6	1.2	155	22	ABG05294	Novel human diagno	810	6	1.2	173	22	AAW38862	Novel subtilisin h
738	6	1.2	155	22	ABW12375	Human bone marrow	811	6	1.2	173	22	AAW38863	Novel subtilisin h
739	6	1.2	155	22	AAW93896	Human protein sequ	812	6	1.2	173	22	AAW38869	Novel subtilisin h

813	6	1.2	173	22	AAU38870	Novel subtilisin h	886	6	1.2	195	23	ABU52193	Helicobacter pylor
814	6	1.2	173	22	AAU38872	Novel subtilisin h	887	6	1.2	195	23	ABP66916	Human polypeptide
815	6	1.2	174	21	AAU57727	Arabidopsis thalia	888	6	1.2	195	23	ABP47916	Human polypeptide
816	6	1.2	174	22	ABG19365	Novel human diago	889	6	1.2	196	21	AAU39749	Arabidopsis thalia
817	6	1.2	174	22	ABG28588	Novel human diago	890	6	1.2	196	21	AAU39749	Arabidopsis thalia
818	6	1.2	174	23	ABG78123	ITLV, LOR-2, STRI	891	6	1.2	196	21	AAU39749	Arabidopsis thalia
819	6	1.2	174	24	ABG84667	Human SECP-19 prot	892	6	1.2	196	21	AAU39749	Arabidopsis thalia
820	6	1.2	175	17	ABG90793	K3 region reverse	893	6	1.2	196	21	AAU39749	Arabidopsis thalia
821	6	1.2	175	22	AAH85126	C. parvum cryptopa	894	6	1.2	196	21	AAU39749	Arabidopsis thalia
822	6	1.2	175	23	ABJ04062	ICP34.5 fragment	895	6	1.2	197	23	ABG72351	Soybean partial py
823	6	1.2	177	23	ABG78049	ITLV, LOR-2, STRI	896	6	1.2	198	22	AAU50280	Streptococcus poly
824	6	1.2	177	23	ABP28875	Streptococcus poly	897	6	1.2	198	22	AAU50280	Streptococcus poly
825	6	1.2	178	20	AAU48389	Human prostate can	898	6	1.2	199	14	AAU34398	Helicobacter pylor
826	6	1.2	178	21	AAU27658	Arabidopsis thalia	899	6	1.2	199	17	AAU34398	Helicobacter pylor
827	6	1.2	178	22	ABG18621	Novel human diago	900	6	1.2	199	17	AAU34398	Helicobacter pylor
828	6	1.2	179	14	AAU50056	ICP34.5 fragment	901	6	1.2	199	22	AAU35692	Helicobacter pylor
829	6	1.2	179	23	ABP29015	Streptococcus poly	902	6	1.2	199	22	AAU35692	Helicobacter pylor
830	6	1.2	180	19	AAU69665	Human NY-ESO-1 pro	903	6	1.2	200	23	AAU35692	Helicobacter pylor
831	6	1.2	180	19	AAU69665	Human NY-ESO-1 pro	904	6	1.2	200	23	AAU35692	Helicobacter pylor
832	6	1.2	180	20	AAU69665	Human NY-ESO-1 pro	905	6	1.2	201	21	AAU35692	Helicobacter pylor
833	6	1.2	180	21	AAU69665	Human NY-ESO-1 pro	906	6	1.2	201	21	AAU35692	Helicobacter pylor
834	6	1.2	180	21	AAU69665	Human NY-ESO-1 pro	907	6	1.2	201	21	AAU35692	Helicobacter pylor
835	6	1.2	180	21	AAU69665	Human NY-ESO-1 pro	908	6	1.2	201	21	AAU35692	Helicobacter pylor
836	6	1.2	180	21	AAU69665	Human NY-ESO-1 pro	909	6	1.2	201	21	AAU35692	Helicobacter pylor
837	6	1.2	180	21	AAU69665	Human NY-ESO-1 pro	910	6	1.2	201	21	AAU35692	Helicobacter pylor
838	6	1.2	180	22	AAU69665	Human NY-ESO-1 pro	911	6	1.2	201	21	AAU35692	Helicobacter pylor
839	6	1.2	180	22	AAU69665	Human NY-ESO-1 pro	912	6	1.2	201	21	AAU35692	Helicobacter pylor
840	6	1.2	180	22	AAU69665	Human NY-ESO-1 pro	913	6	1.2	201	21	AAU35692	Helicobacter pylor
841	6	1.2	180	22	AAU69665	Human NY-ESO-1 pro	914	6	1.2	201	21	AAU35692	Helicobacter pylor
842	6	1.2	180	22	AAU69665	Human NY-ESO-1 pro	915	6	1.2	201	21	AAU35692	Helicobacter pylor
843	6	1.2	180	23	AAU69665	Human NY-ESO-1 pro	916	6	1.2	201	21	AAU35692	Helicobacter pylor
844	6	1.2	180	23	AAU69665	Human NY-ESO-1 pro	917	6	1.2	201	21	AAU35692	Helicobacter pylor
845	6	1.2	180	24	AAU69665	Human NY-ESO-1 pro	918	6	1.2	201	21	AAU35692	Helicobacter pylor
846	6	1.2	180	24	AAU69665	Human NY-ESO-1 pro	919	6	1.2	201	21	AAU35692	Helicobacter pylor
847	6	1.2	180	24	AAU69665	Human NY-ESO-1 pro	920	6	1.2	201	21	AAU35692	Helicobacter pylor
848	6	1.2	180	24	AAU69665	Human NY-ESO-1 pro	921	6	1.2	201	21	AAU35692	Helicobacter pylor
849	6	1.2	180	24	AAU69665	Human NY-ESO-1 pro	922	6	1.2	201	21	AAU35692	Helicobacter pylor
850	6	1.2	181	20	AAU69665	Human NY-ESO-1 pro	923	6	1.2	201	21	AAU35692	Helicobacter pylor
851	6	1.2	181	22	AAU69665	Human NY-ESO-1 pro	924	6	1.2	201	21	AAU35692	Helicobacter pylor
852	6	1.2	181	22	AAU69665	Human NY-ESO-1 pro	925	6	1.2	201	21	AAU35692	Helicobacter pylor
853	6	1.2	181	24	AAU69665	Human NY-ESO-1 pro	926	6	1.2	201	21	AAU35692	Helicobacter pylor
854	6	1.2	184	21	AAU69665	Human NY-ESO-1 pro	927	6	1.2	201	21	AAU35692	Helicobacter pylor
855	6	1.2	184	22	AAU69665	Human NY-ESO-1 pro	928	6	1.2	201	21	AAU35692	Helicobacter pylor
856	6	1.2	185	23	AAU69665	Human NY-ESO-1 pro	929	6	1.2	201	21	AAU35692	Helicobacter pylor
857	6	1.2	185	23	AAU69665	Human NY-ESO-1 pro	930	6	1.2	201	21	AAU35692	Helicobacter pylor
858	6	1.2	187	22	AAU69665	Human NY-ESO-1 pro	931	6	1.2	201	21	AAU35692	Helicobacter pylor
859	6	1.2	187	24	AAU69665	Human NY-ESO-1 pro	932	6	1.2	201	21	AAU35692	Helicobacter pylor
860	6	1.2	187	24	AAU69665	Human NY-ESO-1 pro	933	6	1.2	201	21	AAU35692	Helicobacter pylor
861	6	1.2	188	19	AAU69665	Human NY-ESO-1 pro	934	6	1.2	201	21	AAU35692	Helicobacter pylor
862	6	1.2	188	24	AAU69665	Human NY-ESO-1 pro	935	6	1.2	201	21	AAU35692	Helicobacter pylor
863	6	1.2	189	19	AAU69665	Human NY-ESO-1 pro	936	6	1.2	201	21	AAU35692	Helicobacter pylor
864	6	1.2	189	21	AAU69665	Human NY-ESO-1 pro	937	6	1.2	201	21	AAU35692	Helicobacter pylor
865	6	1.2	189	24	AAU69665	Human NY-ESO-1 pro	938	6	1.2	201	21	AAU35692	Helicobacter pylor
866	6	1.2	190	21	AAU69665	Human NY-ESO-1 pro	939	6	1.2	201	21	AAU35692	Helicobacter pylor
867	6	1.2	190	22	AAU69665	Human NY-ESO-1 pro	940	6	1.2	201	21	AAU35692	Helicobacter pylor
868	6	1.2	190	22	AAU69665	Human NY-ESO-1 pro	941	6	1.2	201	21	AAU35692	Helicobacter pylor
869	6	1.2	190	23	AAU69665	Human NY-ESO-1 pro	942	6	1.2	201	21	AAU35692	Helicobacter pylor
870	6	1.2	191	17	AAU69665	Human NY-ESO-1 pro	943	6	1.2	201	21	AAU35692	Helicobacter pylor
871	6	1.2	191	20	AAU69665	Human NY-ESO-1 pro	944	6	1.2	201	21	AAU35692	Helicobacter pylor
872	6	1.2	191	21	AAU69665	Human NY-ESO-1 pro	945	6	1.2	201	21	AAU35692	Helicobacter pylor
873	6	1.2	191	21	AAU69665	Human NY-ESO-1 pro	946	6	1.2	201	21	AAU35692	Helicobacter pylor
874	6	1.2	191	22	AAU69665	Human NY-ESO-1 pro	947	6	1.2	201	21	AAU35692	Helicobacter pylor
875	6	1.2	191	22	AAU69665	Human NY-ESO-1 pro	948	6	1.2	201	21	AAU35692	Helicobacter pylor
876	6	1.2	191	22	AAU69665	Human NY-ESO-1 pro	949	6	1.2	201	21	AAU35692	Helicobacter pylor
877	6	1.2	192	22	AAU69665	Human NY-ESO-1 pro	950	6	1.2	201	21	AAU35692	Helicobacter pylor
878	6	1.2	193	22	AAU69665	Human NY-ESO-1 pro	951	6	1.2	201	21	AAU35692	Helicobacter pylor
879	6	1.2	194	22	AAU69665	Human NY-ESO-1 pro	952	6	1.2	201	21	AAU35692	Helicobacter pylor
880	6	1.2	194	23	AAU69665	Human NY-ESO-1 pro	953	6	1.2	201	21	AAU35692	Helicobacter pylor
881	6	1.2	195	22	AAU69665	Human NY-ESO-1 pro	954	6	1.2	201	21	AAU35692	Helicobacter pylor
882	6	1.2	195	22	AAU69665	Human NY-ESO-1 pro	955	6	1.2	201	21	AAU35692	Helicobacter pylor
883	6	1.2	195	22	AAU69665	Human NY-ESO-1 pro	956	6	1.2	201	21	AAU35692	Helicobacter pylor
884	6	1.2	195	22	AAU69665	Human NY-ESO-1 pro	957	6	1.2	201	21	AAU35692	Helicobacter pylor
885	6	1.2	195	22	AAU69665	Human NY-ESO-1 pro	958	6	1.2	201	21	AAU35692	Helicobacter pylor

```
959 6 1.2 214 22 AAU17093 Novel signal trans
960 6 1.2 214 23 AAO18339 Mature humanised m
961 6 1.2 215 21 AAG19526 Arabidopsis thalia
962 6 1.2 216 21 AAG57236 Arabidopsis thalia
963 6 1.2 216 21 AAG61182 Arabidopsis thalia
964 6 1.2 216 22 ABBS5033 Streptococcus pyos
965 6 1.2 217 21 AAU71042 Streptococcus pyos
966 6 1.2 217 22 ABBS0591 Streptococcus pyos
967 6 1.2 217 22 AAU41274 Propionibacterium
968 6 1.2 217 22 AAU42502 Streptococcus poly
969 6 1.2 217 23 ABP30013 Streptococcus poly
970 6 1.2 218 21 AAB16343 Plms radata alph
971 6 1.2 218 22 ABB70993 Droscophila melano
972 6 1.2 219 21 AAY76981 Quail H1 histone p
973 6 1.2 219 23 ABP29022 Streptococcus poly
974 6 1.2 222 20 AAY36900 Chlamydia trachom
975 6 1.2 222 22 ABG08919 Novel human diagno
976 6 1.2 222 22 AAU17547 Novel signal trans
977 6 1.2 223 22 AAU30182 Novel human secret
978 6 1.2 224 20 AAM96229 Sma-4 protein C-te
979 6 1.2 224 21 AAY87590 Aminoamidase pre
980 6 1.2 224 22 ABG01822 Novel human diagno
981 6 1.2 224 23 ABB81607 A. thaliana glucos
982 6 1.2 225 22 AAU35452 Haemophilus influe
983 6 1.2 225 22 ABG02862 Novel human diagno
984 6 1.2 225 22 ABG03051 Novel human diagno
985 6 1.2 225 22 AAU27631 Human protein AFp2
986 6 1.2 225 22 AAU17137 Novel signal trans
987 6 1.2 226 21 AAG20903 Arabidopsis thalia
988 6 1.2 226 21 AAG37582 Arabidopsis thalia
989 6 1.2 226 22 ABB83906 Droscophila melano
990 6 1.2 226 22 AAB83964 Partial amino acid
991 6 1.2 226 23 ABB81606 A. thaliana glucos
992 6 1.2 227 22 ABG16125 Novel human diagno
993 6 1.2 228 23 ABUS1752 Helicobacter pylor
994 6 1.2 229 20 AAY29178 Amino acid sequenc
995 6 1.2 229 21 AAG41580 Arabidopsis thalia
996 6 1.2 229 23 ABG66716 Human novel polype
997 6 1.2 230 21 AAG29871 Arabidopsis thalia
998 6 1.2 233 16 AAB67918 (1-3)-beta-D-gluc
999 6 1.2 233 23 ABP65358 Bifidobacterium 10
1000 6 1.2 233 23 ABP65603 Bifidobacterium 10
```

ALIGNMENTS

```
RESULT 1
AAB67668
ID AAB67668 standard; Protein; 502 AA.
XX
AC AAB67668;
XX
DT 11-JUN-2001 (first entry)
XX
DE Amino acid sequence of a BASB109 polypeptide.
XX
KM BASB109; bacterial infection; vaccine; genetic immunisation.
XX
OS Moraxella catarrhalis.
XX
PN MO200119996-A1.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-EF09035.
XX
PR 14-SEP-1999; 99GB-0021691.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
```

```
DR MPI; 2001-257883/26.
DR N-PSDB; AAF55658.
XX
PT Novel BASB109 polypeptides of Moraxella catarrhalis useful for
PT diagnostic, prophylactic and therapeutic purposes against microbial
PT diseases, preferably bacterial infections
XX
PS Claim 3; Page 65; 93pp; English.
XX
CC The present sequence represents a BASB109 polypeptide of Moraxella
CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for
CC treating bacterial infections, and as research reagents and materials
CC for the treatment of and diagnosis of diseases, particularly human
CC diseases. They are useful for inducing an immune response in an
CC individual, and to assess the binding of small molecule substrates and
CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,
CC and natural product mixtures. BASB109 polynucleotides are useful for
CC therapeutic or prophylactic purposes, in particular genetic immunisation
CC and in diagnosis of the stage and type of infection.
XX
SQ Sequence 502 AA;
Query Match 100.0%; Score 502; DB 22; Length 502;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MSKPTLIKTTLCALSAALMISGCSNOADKAQPKSTVDAAKTANADNAASOEHOGEIP 60
DB 1 MSKPTLIKTTLCALSAALMISGCSNOADKAQPKSTVDAAKTANADNAASOEHOGEIP 60
QY 61 VIDAIYTHAPEVPPVDRDHPAKVVKMETVEKVKRLADGVEYQFTFGGQVPGQWIRVR 120
DB 61 VIDAIYTHAPEVPPVDRDHPAKVVKMETVEKVKRLADGVEYQFTFGGQVPGQWIRVR 120
QY 121 EGGTIVQFSNHPDSMPHNVDPHATGPGGAESAFTAPGHTSTFSFALQPLGVYHC 180
DB 121 EGGTIVQFSNHPDSMPHNVDPHATGPGGAESAFTAPGHTSTFSFALQPLGVYHC 180
QY 121 EGGTIVQFSNHPDSMPHNVDPHATGPGGAESAFTAPGHTSTFSFALQPLGVYHC 180
DB 121 EGGTIVQFSNHPDSMPHNVDPHATGPGGAESAFTAPGHTSTFSFALQPLGVYHC 180
QY 181 AAVPVMHIANMGYGLILVEPKEGLPKVDKEYVWQGDYTTKRGYGEGLQPPDEKAIR 240
DB 181 AAVPVMHIANMGYGLILVEPKEGLPKVDKEYVWQGDYTTKRGYGEGLQPPDEKAIR 240
QY 241 EDAEYVVFNGSVGALTGENALKAQGETVRLFVNGGPNLTSFHVIGEIFDKVHEGK 300
DB 241 EDAEYVVFNGSVGALTGENALKAQGETVRLFVNGGPNLTSFHVIGEIFDKVHEGK 300
QY 301 GENHNIQTLLIPAGGAITEPKVDVPGDVVYVDHAIFFAFNKGALGLVVEGEENHEIYS 360
DB 301 GENHNIQTLLIPAGGAITEPKVDVPGDVVYVDHAIFFAFNKGALGLVVEGEENHEIYS 360
QY 361 HKOTDAVYLPPEGAPQAIIDQEAAPKTPAPANLQEOIKAGATYDSNCAACHOPDGKVPNA 420
DB 361 HKOTDAVYLPPEGAPQAIIDQEAAPKTPAPANLQEOIKAGATYDSNCAACHOPDGKVPNA 420
QY 421 FPEPLANSDYLNADHARAASIVANGLSGITVNGNOYESVMPAIALSDQOIANVITYTLNS 480
DB 421 FPEPLANSDYLNADHARAASIVANGLSGITVNGNOYESVMPAIALSDQOIANVITYTLNS 480
QY 481 FGNKGQLSADVAAYAKKTKPN 502
DB 481 FGNKGQLSADVAAYAKKTKPN 502
XX
RESULT 2
AAB67669
ID AAB67669 standard; Protein; 502 AA.
XX
AC AAB67669;
XX
DT 11-JUN-2001 (first entry)
XX
DE Amino acid sequence of a BASB109 polypeptide.
XX
```

KM BASB109; bacterial infection; vaccine; genetic immunisation.
 XX Moraxella catarrhalis.
 OS WQ200119996-A1.
 PN
 XX
 PD 22-MAR-2001.
 XX
 XX 14-SEP-2000; 2000WO-EP09035.
 XX
 XX 14-SEP-1999; 99GB-0021691.
 PR
 XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA
 XX Thonnard J;
 XX
 XX WPI: 2001-257883/26.
 DR N-PSDB; AAF55657.
 DR
 XX Novel BASB109 polypeptides of Moraxella catarrhalis useful for
 PT diagnostic, prophylactic and therapeutic purposes against microbial
 PT diseases, preferably bacterial infections
 XX
 PS Claim 3; Page 66; 93pp; English.
 XX
 CC The present sequence represents a BASB109 polypeptide of Moraxella
 CC catarrhalis. BASB109 polypeptides and polynucleotides are useful for
 CC treating bacterial infections, and as research reagents and materials
 CC for the treatment of and diagnosis of diseases, particularly human
 CC diseases. They are useful for inducing an immune response in an
 CC individual, and to assess the binding of small molecule substrates and
 CC ligands in, for e.g. cells, cell-free preparations, chemical libraries,
 CC and natural product mixtures. BASB109 polynucleotides are useful for
 CC therapeutic or prophylactic purposes, in particular genetic immunisation
 CC and in diagnosis of the stage and type of infection.
 CC
 SQ Sequence 502 AA;
 Query Match 100.0%; Score 502; DB 22; Length 502;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKPTLITKTLTICALSALMLSCGCSNQADRAQPKSSTVDAAKTANADNAAQEHGELP 60
 DB 1 MSKPTLITKTLTICALSALMLSCGCSNQADRAQPKSSTVDAAKTANADNAAQEHGELP 60
 QY 61 VIDAIVTAPEVPPVDRDHPAKVYVYKMEVEMKLAGVYQFTFGQVPGQMITRVR 120
 DB 61 VIDAIVTAPEVPPVDRDHPAKVYVYKMEVEMKLAGVYQFTFGQVPGQMITRVR 120
 QY 121 EDDTIEVQSNHPDSKPHNVDPFAATGPGGGAESAFTAPGHTSTPSFKALQGLVYVHC 180
 DB 121 EDDTIEVQSNHPDSKPHNVDPFAATGPGGGAESAFTAPGHTSTPSFKALQGLVYVHC 180
 QY 181 AVAPVGMHIANGYGLIVEPEKGLPKVKEYVVMQDGFYTKGKYEQGLQPFDMKAR 240
 DB 181 AVAPVGMHIANGYGLIVEPEKGLPKVKEYVVMQDGFYTKGKYEQGLQPFDMKAR 240
 QY 241 EDAEYVVFNGSVCALTGENAALKAKVGETVRLFPVNGGPNUTSSFHYIGTIFDKVHEGK 300
 DB 241 EDAEYVVFNGSVCALTGENAALKAKVGETVRLFPVNGGPNUTSSFHYIGTIFDKVHEGK 300
 QY 301 GENNNIQTLLIPAGGAITEFKVDVGVYLVDAHFAFNKCALILKVEGENHEIYS 360
 DB 301 GENNNIQTLLIPAGGAITEFKVDVGVYLVDAHFAFNKCALILKVEGENHEIYS 360
 QY 361 HKQTDVYVLPFGAPQAIIDQAPKTPAPANLQEQIRAGATYDSNCAACHQPDGKGVNA 420
 DB 361 HKQTDVYVLPFGAPQAIIDQAPKTPAPANLQEQIRAGATYDSNCAACHQPDGKGVNA 420
 QY 421 FPLPLNSDVLNADHAAASIVANGLSGKTTNGNÖYESVMPAIALSDOOIANTVYTLNS 480
 DB 421 FPLPLNSDVLNADHAAASIVANGLSGKTTNGNÖYESVMPAIALSDOOIANTVYTLNS 480

QY 481 FGNKGQSLSDVDVAKAKTKTPN 502
 DB 481 FGNKGQSLSDVDVAKAKTKTPN 502
 RESULT 3
 AAO17561
 ID AAO17561 standard; Protein; 197 AA.
 XX
 AC AAO17561;
 XX
 DT 19-UTL-2002 (first entry)
 XX
 DE M catarrhalis MCA100170 protein SEQ ID NO: 2.
 XX
 XX Moraxella; vaccine; respiratory tract infection; antiinflammatory;
 KW auditory; antibacterial; otitis media; sinusitis; pneumonia.
 OS Moraxella catarrhalis.
 XX
 PN WQ200218595-A2.
 XX
 PD 07-MAR-2002.
 XX
 XX 28-AUG-2001; 2001WO-CA01221.
 XX
 PR 28-AUG-2000; 2000US-228294P.
 PR 28-AUG-2000; 2000US-228295P.
 PR 28-AUG-2000; 2000US-228296P.
 PR 29-AUG-2000; 2000US-228438P.
 PR 29-AUG-2000; 2000US-228439P.
 PR 29-AUG-2000; 2000US-228440P.
 PR 29-AUG-2000; 2000US-228441P.
 PR 29-AUG-2000; 2000US-228442P.
 PR 29-AUG-2000; 2000US-228443P.
 PR 29-AUG-2000; 2000US-228511P.
 PR 29-AUG-2000; 2000US-228512P.
 PR 29-AUG-2000; 2000US-228742P.
 PR 29-AUG-2000; 2000US-228773P.
 PR 01-SEP-2000; 2000US-229465P.
 PR 01-SEP-2000; 2000US-229474P.
 PR 01-SEP-2000; 2000US-229475P.
 PR 01-SEP-2000; 2000US-229478P.
 PR 05-SEP-2000; 2000US-229740P.
 PR 05-SEP-2000; 2000US-229803P.
 PR 05-SEP-2000; 2000US-229804P.
 PR 05-SEP-2000; 2000US-229805P.
 PR 05-SEP-2000; 2000US-229806P.
 PR 05-SEP-2000; 2000US-229809P.
 PR 05-SEP-2000; 2000US-229811P.
 PR 06-SEP-2000; 2000US-230214P.
 PR 06-SEP-2000; 2000US-230250P.
 PR 06-SEP-2000; 2000US-230252P.
 PA (AVET) AVENTIS PASTEUR LTD.
 XX
 PI Loesmore S, Wang J, Bradley B, Ochs M, Yang Y;
 DR WPI: 2002-401721/43.
 DR N-PSDB; AAL46493.
 XX
 PT Moraxella polypeptide and polynucleotides useful as vaccine for
 PT immunizing a host e.g. humans against disease e.g. otitis media,
 PT pneumonia, caused by infection of the bacteria
 XX
 PS Claim 28; Fig 1; 277pp; English.
 CC The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a protein of the invention.

XX Sequence 197 AA;
 SQ Query Match 39.2%; Score 197; DB 23; Length 197;
 Best Local Similarity 100.0%; Pred. No. 7.4e-183;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKPTLTKTLTLCALSLMLSGCSNOADKKAQPKSSTVDAAKATNADNMAOEHQGLP 60
 DB 1 MSKPTLTKTLTLCALSLMLSGCSNOADKKAQPKSSTVDAAKATNADNMAOEHQGLP 60
 QY 61 VIDAIVTTHAPVPPVDRDHPAKVYVVKETVKWRLADGVYQFTFGQVPCQMIRV 120
 DB 61 VIDAIVTTHAPVPPVDRDHPAKVYVVKETVKWRLADGVYQFTFGQVPCQMIRV 120
 QY 121 EGDITLVOFSNHPDSKMPHNVDFAATGPGGAEASFTAPGHTSFSPKALQPLYYVHC 180
 DB 121 EGDITLVOFSNHPDSKMPHNVDFAATGPGGAEASFTAPGHTSFSPKALQPLYYVHC 180
 QY 181 AVAPVGMHIANGMVGLI 197
 DB 181 AVAPVGMHIANGMVGLI 197

RESULT 4
 ABP77627 ID ABP77627 standard; Protein; 392 AA.
 XX ABP77627;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 1784.
 XX KM Antibacterial; infection; vaccine; gene therapy.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-IB02069.
 XX PR 12-FEB-2001; 2001GB-0003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Piazza M, Maignani V, Monaci E;
 XX DR WPI; 2003-058415/05.
 XX DR N-PSDB; AB238597.
 XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX PT medicament for treating or preventing N. gonorrhoeae infection -
 XX PS Disclosure; Page 316; 815pp; English.
 XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
 XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
 XX CC antibodies that specifically bind to the proteins. The composition
 XX CC comprising the protein, nucleic acid or antibody is useful for the
 XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 XX CC infection, this may be in the form of a vaccine or gene therapy.
 XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 XX CC molecules of the invention.
 XX SQ Sequence 392 AA;

Query Match 6.8%; Score 34; DB 24; Length 392;
 Best Local Similarity 100.0%; Pred. No. 3.4e-24;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 YHCAPVGMHIANGMVGLIVPEKGLPKVDKE 211
 DB 173 YHCAPVGMHIANGMVGLIVPEKGLPKVDKE 206

RESULT 5
 AAY05975 ID AAY05975 standard; Peptide; 10 AA.
 XX AC AAY05975;
 XX DT 16-AUG-1999 (first entry)
 XX DE Human cancer antigen NY ESO-1/CAG-3 ORF1 cancer peptide.
 XX KM NY ESO-1/CAG-3 gene; CAG-3 gene; cancer peptide; antigen; human;
 XX KM leukemia; non-Hodgkins lymphoma; Hodgkins lymphoma; lung cancer;
 XX KM metastasis; melanoma; adenocarcinoma; thymoma; colon cancer;
 XX KM uterine cancer; breast cancer; prostate cancer; ovarian cancer;
 XX KM cervical cancer; bladder cancer; kidney cancer; pancreatic cancer;
 XX KM liver cancer; sarcoma; tumour; diagnosis; immunotherapy; therapy;
 XX KM vaccine.
 XX OS Homo sapiens.
 XX OS Synthetic.
 XX PN WO9918206-A2.
 XX PD 15-APR-1999.
 XX PF 21-SEP-1998; 98WO-US19609.
 XX PR 08-OCT-1997; 97US-0061428.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Rosenberg SA, Wang RF;
 XX DR WPI; 1999-277270/23.
 XX PT Cancer antigen NY ESO1/CAG-3
 XX PS Claim 29; Page 10; 88pp; English.
 XX CC The present sequence represents a cancer peptide that is based on
 XX CC amino acid residues 55-62 of human ESO-1/CAG-3 (or CAG-3) ORF1
 XX CC (see AAY05965), a new and potent tumour antigen capable of eliciting
 XX CC an antigen specific immune response by T cells. Cancer peptides
 XX CC derived from CAG-3 ORF1, CAG-3 ORF2 (see AAY05966), portions of them
 XX CC and their variants (see AAY05967-87), are useful as cancer vaccines
 XX CC that protect against cancer. The invention provides: vectors and
 XX CC host cells (also useful as vaccines); a method of diagnosis of
 XX CC cancer or precancer; a transgenic animal; antisense oligonucleotides
 XX CC that inhibit expression of the cancer peptide or tumour antigen;
 XX CC antibodies reacting with a CAG-3 cancer peptide, useful in
 XX CC diagnostic and detection assays; and methods for preventing or
 XX CC inhibiting cancer by administering a cancer peptide, with or without
 XX CC an HLA molecule. The cancer peptides form part of, or are derived
 XX CC from, cancers such as primary or metastatic melanoma, thymoma,
 XX CC lymphoma, sarcoma, lung cancer, liver cancer, leukemia, uterine
 XX CC cancer, cervical cancer, bladder cancer, kidney cancer and
 XX CC adenocarcinomas such as breast, prostate, ovarian, pancreatic and
 XX CC thyroid cancer. Melanoma is treated by inducing cancer-specific T
 XX CC cells in vitro for subsequent return to a patient.

QY 146 ATGPGGA 153
 Query Match 1.6%; Score 8; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ATPGCGA 8

RESULT 6
ABG18127

ID ABG18127 standard; Protein; 415 AA.

XX ABG18127;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #18118.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS82314.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 48486; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostic, forensic, gene mapping, identification of mutations in
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 415 AA;

Query Match 1.6%; Score 8; DB 22; Length 415;

Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 HARRASIV 441

Db 250 HARRASIV 257

RESULT 7
AAV74923

ID AAV74923 standard; Protein; 497 AA.

XX AAV74923;

DT 21-MAR-2000 (first entry)

DE Neisseria gonorrhoeae ORF 402 protein sequence SEQ ID NO:11320.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.

OS Neisseria gonorrhoeae.

PN WO9957280-A2.

PD 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.

PI Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappelli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;

DR WPI; 2000-062150/05.

DR N-PSDB; AAZ53685.

PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -

XX Claim 2; Page 724; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAV74253 to AAV75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.

SQ Sequence 497 AA;

Query Match 1.6%; Score 8; DB 21; Length 497;

Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SSTVDAAA 42

Db 454 SSTVDAAA 461

RESULT 8

```

AAV74924
ID AAV74924 standard; Protein; 497 AA.
XX
XX AAV74924;
AC
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 402 protein sequence SEQ ID NO:1322.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
OS
XX
XX WO957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
XX
XX N-PSDB; AA253686.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 2; Page 725; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX
SQ Sequence 497 AA;

Query Match 1.6%; Score 8; DB 21; Length 497;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SSTVDAAA 42
DB 454 SSTVDAAA 461

```

```

XX
XX AAV74925;
AC
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 402 protein sequence SEQ ID NO:1324.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
OS
XX
XX WO957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
XX
XX N-PSDB; AA253687.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 2; Page 726-727; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX
SQ Sequence 497 AA;

Query Match 1.6%; Score 8; DB 21; Length 497;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SSTVDAAA 42
DB 454 SSTVDAAA 461

```

RESULT 9
AAV74925
ID AAV74925 standard; Protein; 497 AA.

RESULT 10
ABP80199
ID ABP80199 standard; Protein; 508 AA.
XX
XX ABP80199;

CC overexpress it, or express it in a non-toxic mutant form.
XX
SQ Sequence 697 AA;

Query Match 1.6%; Score 8; DB 21; Length 697;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SSTVDAAA 42
|||
Db 654 SSTVDAAA 661

RESULT 13
ABP65660
ID ABP65660 standard; Protein; 995 AA.

AC ABP65660;

DT 19-NOV-2002 (first entry)

DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:404.

XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KM anti-diarrheic; antibacterial; inhibitor of Salmonella; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX rotavirus; food composition; pharmaceutical composition.

OS Bifidobacterium longum.

PN EPI227152-A1.

PD 31-JUL-2002.

PF 30-JAN-2001; 2001EP-0102050.

PR 30-JAN-2001; 2001EP-0102050.

XX (NEST) SOC PROD NESTLE SA.

XX WPI; 2002-668397/72.

PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
as a probe or primer for detecting and/or identifying Bifidobacterium
longum in a biological sample -

PS Claim 3; SEQ ID 404; 80bp; English.

XX The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in AB081842 and AB081843, or a sequence exhibiting at
CC least 90% identity or which hybridizes with the sequences given in
CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding
CC a fusion protein, comprising a sequence selected from 1097 sequences
CC given in AB655258 to ABP6534 ligated in frame to a polynucleotide
CC encoding a heterologous polypeptide. (I) has anti-diarrheic and
CC antibacterial activities, and can be used as an inhibitor of Salmonella.
CC (I) (which is a probe) is useful for the detection and/or identification
CC of Bifidobacterium longum in a biological sample. A carrier containing
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618)
CC can be used for preventing and/or treating diarrhoea brought about by
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC fermented products, ice-creams, fermented cereal based products, milk
CC based powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC listing from the present invention but not mentioned further within the
CC N.B. The sequence data for this patent is not represented in the printed

CC specification but is based on sequence information supplied by the
CC European Patent Office.
XX

SQ Sequence 995 AA;

Query Match 1.6%; Score 8; DB 23; Length 995;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 489 SADDVAKA 496
|||
Db 40 SADDVAKA 47

RESULT 14
AAB12246
ID AAB12246 standard; peptide; 33 AA.

AC AAB12246;

DT 10-NOV-2000 (first entry)

DE Partial sequence of HIV-1 strain DUR gp41 descending helix.

XX HIV-1; AIDS; human immunodeficiency virus type 1; group O HIV;
KM acquired immunodeficiency syndrome; group M HIV; gp41; DUR.
XX Human immunodeficiency virus type 1.

OS Human immunodeficiency virus type 1.

PN EPI013766-A2.

PD 28-JUN-2000.

PF 29-NOV-1999; 99EP-0309491.

PR 30-NOV-1998; 98US-0110292.

PR 08-FEB-1999; 99US-0119138.

PR 04-NOV-1999; 99US-0433428.

XX (ORTH-) ORTHO-CLINICAL DIAGNOSTICS INC.

XX De Lays R, Zheng J;

XX WPI; 2000-402205/35.

PT New antigenic peptides and peptide functional derivatives, useful for
PT detection of antibodies produced in response to human immunodeficiency
PT virus group O antibodies -

PS Example 4; Fig 4; 52bp; English.

XX The present sequence is a partial gp41 protein of Human Immunodeficiency
CC virus Type 1 (HIV-1) strain DUR. HIV is the principle aetiological
CC agent for acquired immunodeficiency syndrome (AIDS). gp41 is a HIV
CC envelope protein, and may be used as an antigen for the detection of
CC antibodies produced in response to HIV infection. DUR is a member of
CC HIV group O. The present sequence is the descending helix of gp41. This
CC sequence was used in a sequence homology alignment, which in turn was
CC used to derive consensus sequence peptides (AAB12259 and AAB12260). The
CC peptides of AAB12259 and AAB12260 were used as group O replacement
CC peptides. The peptides of AAB12259 and AAB12260 were used to construct
CC mosaic gp41 proteins, in which the group M immunodominant region was
CC replaced by the peptides of AAB12259 or AAB12260. The mosaic gp41
CC proteins (AAB12261 and AAB12262) would be useful as antigens, which
CC would be used for the detection of anti-group O HIV antibodies produced
CC in response to HIV infection.

SQ Sequence. 33 AA;

Query Match 1.4%; Score 7; DB 21; Length 33;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOQIANV 473
DB 1 DOQIANV 7

RESULT 15

AAAG93319 standard; Protein; 39 AA.

AAAG93319;

13-SEP-2001 (first entry)

Human protein HP10511.

Human; gene therapy; tumour.

Homo sapiens.

WO200142302-A1.

14-JUN-2001.

06-DEC-2000; 2000WO-IP08631.

06-DEC-1999; 99JP-0346863.

06-DEC-1999; 99JP-0346864.

08-FEB-2000; 2000JP-0031062.

10-FEB-2000; 2000JP-0034090.

10-FEB-2000; 2000JP-0034091.

14-FEB-2000; 2000JP-0035829.

14-FEB-2000; 2000JP-0035829.

14-MAR-2000; 2000JP-0071161.

30-MAY-2000; 2000JP-0160851.

(NISC-) JAPAN SCI & TECHNOLOGY CORP.

Kato S, Eguchi C, Saeki M;

WPI; 2001-381646/40.

N-PSDB; AAH68604.

Human protein originated from tumor cell line, applicable as drug,

reagent for studying intracellular protein networks and protein source

for drug screening, also encoded cDNA for gene diagnosis and gene

therapy

Claim 1; Page 416; 471pp; Japanese.

The present sequence is a human protein. The human protein, preferably

originated from tumour cell line, is applicable as a drug, a reagent for

studying intracellular protein networks and a protein source for

screening proteins for binding low molecular weight drugs. The human

protein coding sequence is useful for gene diagnosis and gene therapy,

expression vectors and transformant cells for detection of ligands and

receptors.

Sequence 39 AA;

Query Match 1.4%; Score 7; DB 22; Length 39;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 GGAARS 156

DB 6 GGAARS 12

RESULT 16

ABG55536

ID ABG55536 standard; Peptide; 71 AA.

XX ABG55536;

XX 25-FEB-2003 (first entry)

Human liver peptide, SEQ ID No 34184.

Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

hypercholesterolaemia; coronary heart disease.

Homo sapiens.

MO200157273-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00664.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488898/53.

Human genome-derived single exon nucleic acid probes useful for

analysing gene expression in human adult liver

Claim 27; SEQ ID No 34184; 658pp; English.

The invention relates to a single exon nucleic acid probe (SENP) (I) for

measuring human gene expression in a sample derived from human adult

liver, comprising one of 13109 defined nucleotide sequences given in the

specification (or complements/ fragments). The probe hybridises at high

stringency to a nucleic acid molecule expressed in the human adult

liver. (I) may be used for predicting, measuring and displaying gene

expression in samples derived from human adult liver. The genes

identified may be involved in genetic liver diseases such as cirrhosis,

hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

is associated with coronary heart disease. ABG47348-ABG5930 represent

human liver single exon encoded peptides of the invention.

Note: The sequence information for this patent does not appear in the

printed specification but was obtained in electronic format directly

from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 71 AA;

Query Match 1.4%; Score 7; DB 22; Length 71;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500

DB 28 AKAKTK 34

RESULT 17

ABBA40293

ID ABBA40293 standard; Peptide; 71 AA.

XX ABBA40293;

04-FEB-2002 (first entry)

Peptide #7799 encoded by human foetal liver single exon probe.

Human; foetal liver; gene expression; single exon nucleic acid probe.

XX

OS Homo sapiens.
XX
PN W0200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 32928; 639pp + sequence listing; English.
XX
SQ The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 71 AA;
XX
Query Match 1.4%; Score 7; DB 22; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 494 AKAKTK 500
DB 28 AKAKTK 34
XX
RESULT 18
AAM61086
ID AAM61086 standard; Protein; 71 AA.
XX
AC AAM61086;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33191.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PA W0200157275-A2.
XX
PN 09-AUG-2001.
XX
PD 30-JAN-2001; 2001WO-US00667.
XX
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 33191; 650pp + Sequence listing; English.
XX
SQ The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 71 AA;
XX
Query Match 1.4%; Score 7; DB 22; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 494 AKAKTK 500
DB 28 AKAKTK 34
XX
RESULT 19
AAM73789
ID AAM73789 standard; Protein; 71 AA.
XX
AC AAM73789;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34095.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; Leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PA W0200157276-A2.
XX
PN 09-AUG-2001.
XX
PD 30-JAN-2001; 2001WO-US00668.
XX
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX

PS Example 4; SEQ ID NO: 34095; 658bp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
SQ Sequence 71 AA;

Query Match 1.4%; Score 7; DB 22; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500
|||
Db 28 AKAKTK 34

RESULT 20
AAM33976
ID AAM33976 standard; Protein; 71 AA.

AC AAM33976;
XX 17-OCT-2001 (first entry)

DE Peptide #8013 encoded by probe for measuring placental gene expression.

KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.

OS Homo sapiens.

PN WO200157272-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

PS Claim 27; SEQ ID No 34245; 654bp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI13135-AI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
SQ Sequence 71 AA;

Query Match 1.4%; Score 7; DB 22; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500
|||
Db 28 AKAKTK 34

RESULT 21
ABG43675
ID ABG43675 standard; Peptide; 71 AA.

XX ABG43675;

DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 33340.

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostasis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.

OS Homo sapiens.

PN WO200186003-A2.

XX 15-NOV-2001.

PF 30-JAN-2001; 2001WO-US00665.

PR 04-FEB-2000; 2000US-180312P.

PR 26-MAY-2000; 2000US-207456P.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

PT Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -

PS Claim 27; SEQ ID No 33340; 634bp; English.

XX CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridize at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acid derived from human lung
CC mRNA, and (b) measuring the labeled nucleic acid bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of

expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Rudrak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

SO Sequence 71 AA;

Query Match 1.4%; Score 7; DB 23; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AKAKTK 500
| | | | |
DB 28 AKAKTK 34

RESULT 22
ID ABB60712 standard; Protein: 90 AA.
AC ABB60712;
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 8928.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PT
XX
XX (PEKE) PE GORP NY.
PA
XX
XX Venter JC, Adams M, Li FWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX N-PSDB; ABL04815.
PT
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX
XX Disclosure; SEQ ID NO 8928; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL161840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX

SO Sequence 90 AA;

Query Match 1.4%; Score 7; DB 22; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 EGDTEV 127
| | | | |
DB 75 EGDTEV 81

RESULT 23
ID AAO10927 standard; Protein: 94 AA.
AC AAO10927;
XX
XX
XX 06-NOV-2001 (first entry)
DT
XX
DE Human polypeptide SEQ ID NO 24819.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX
XX WO200164835-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US04927.
PF
XX
XX 28-FEB-2000; 2000US-0515126.
PR
XX
XX 18-MAY-2000; 2000US-0577409.
PT
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-514838/56.
DR
XX
XX N-PSDB; AAI90858.
PT
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PT
XX
XX Claim 20; SEQ ID NO 24819; 1399pp + Sequence Listing; English.
PS
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
XX Sequence 94 AA;

Query Match 1.4%; Score 7; DB 22; Length 94;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 TGGGGA 153
 DB 26 TGGGGA 32

RESULT 24

AA07245
 ID AA07245 standard; Protein; 104 AA.

AC AA07245;

DT 09-APR-1997 (first entry)

DE HIV-1 group O strain DUR gp41 immunodominant region peptide.

KW Human immunodeficiency virus; subgroup; strain; AIDS; homology; envelop;
 gp120; gp41; seropositive; antibody; primer; probe; group O; group M.

OS Human immunodeficiency virus type 1.

PN WO9612809-A2.

PD 02-MAY-1996.

PF 20-OCT-1995; 95WO-FR01391.

PR 03-MAR-1995; 95FR-0002526.

PR 20-OCT-1994; 94FR-0012554.

PA (INSP) INST PASTEUR.

PI Borman A, Charneau P, Clavel F, Cohen JM, Guetard D;
 P1 Donjon de Saint-Martin J, Montagnier L, Quillent C;

DR WPI; 1996-230610/23.

DR N-PSDB; AAT42241.

PT New antigenic HIV-1 group O strain proteins and related nucleic
 acids - useful in diagnosis, vaccines, therapy etc., of infection by
 PT HIV-1 group O strains VAV or DUR

PS Claim 24; Fig 12; 108pp; French.

CC The invention relates to the isolation of a novel subgroup of the human
 CC immunodeficiency virus (HIV) type 1, designated group O. In particular,
 CC the inventors have isolated 2 new strains of the group O virus: strains
 CC VAV and DUR. Strain VAV was isolated from a French AIDS patient and has
 CC homology to the recently characterised Cameroonian HIV strains ANT70 and
 CC WPI5180. The DUR strain was isolated from a seropositive patient from
 CC the Camerouns who showed atypical seroreactivity. The sequence presented
 CC here represents the amino acid sequence of the DUR strain gp41
 CC immunodominant region. The DNA and protein sequences are used to
 CC generate peptides for detection of antibodies from patients infected
 CC with the new group O strains, as well as primers and probes to detect
 CC the viral nucleic acids. The peptides and nucleic acid sequences derived
 CC from these strains are able to distinguish between the group O and group
 CC M viral strains.

SQ Sequence 104 AA;

Query Match 1.4%; Score 7; DB 17; Length 104;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 DOOIANV 473
 DB 83 DOOIANV 89

RESULT 25

ABG26287
 ID ABG26287 standard; Protein; 105 AA.

AC ABG26287;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26278.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-APR-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS90474.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 20; SEQ ID No 5646; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 105 AA;

Query Match 1.4%; Score 7; DB 22; Length 105;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 HARAASI 440
 DB 82 HARAASI 88

RESULT 26

AA72394
ID AAR72394 standard; Protein: 111 AA.
XX
XX
AC AAR72394;
XX
XX
DT 14-DEC-1995 (first entry)
XX
XX
DE Agmenellum quadruplicatum PR-6 rubisco small subunit.
XX
XX
KM Rubisco; ribulose 1,5-bisphosphate carboxylase/oxygenase;
KW large subunit; carbon dioxide fixation; Synechococcus sp.
XX
OS Agmenellum quadruplicatum PR-6.
PN JP07079782-A.
XX
XX
PD 28-MAR-1995.
XX
XX
PF 18-JUN-1993; 93JP-0184304.
XX
PR 18-JUN-1993; 93JP-0184304.
XX
PA (KANT) KANSAI DENRYOKU KK.
XX
XX
DR WPI; 1995-157852/21.
DR N-PSDB; AAQ87824.
XX
XX
PT Promoter sequence of Synechococcus PCC7002-originated rubisco gene -
PT used in a vector to produce Cyano:bacterium with improved carbon
PT d:oxide fixation
XX
XX
PS Claim 12; Page 25; 28pp; Japanese.
XX
CC A sequence including the ORFs coding for the large and small
CC subunits of the rubisco enzyme has been isolated from Agmenellum
CC quadruplicatum PR-6 (Synechococcus sp. strain PCC7002). The
CC region coding for the rubisco small subunit is claimed; the enzyme
CC is involved in carbon dioxide fixation.
XX
XX
SQ Sequence 111 AA;
XX
XX
Query Match 1.4%; Score 7; DB 16; Length 111;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 465 LSDQQA 471
DB 19 LSDQQA 25

RESULT 27
ABB89496
ID ABB89496 standard; Protein: 114 AA.
XX
XX
AC ABB89496;
XX
XX
DT 24-MAY-2002 (first entry)
XX
XX
DE Human polypeptide SEQ ID NO 1872.
XX
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200190304-A2.
XX
XX
PD 29-NOV-2001.
XX
XX
PF 18-MAY-2001; 2001WO-US16450.

XX
XX
XX 19-MAY-2000; 2000US-205515P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Birse CE, Rosen CA;
XX
XX
DR WPI; 2002-122018/16.
DR N-PSDB; ABL99905.
XX
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
XX
PS Claim 11; SEQ ID NO 1872; 2081pp + Sequence Listing; English.
XX
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 114 AA;
XX
XX
Query Match 1.4%; Score 7; DB 23; Length 114;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 ALMLSGC 23
DB 57 ALMLSGC 63

RESULT 28
AAB27869
ID AAB27869 standard; Protein: 118 AA.
XX
XX
AC AAB27869;
XX
XX
DT 29-JAN-2001 (first entry)
XX
XX
DE Sequence homologous to protein fragment encoded by gene 27.
XX
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200055199-A1.
XX
XX
PD 21-SEP-2000.
XX
XX
PF 09-MAR-2000; 2000WO-US06014.
XX
XX
PR 12-MAR-1999; 99US-0124095.
PR 11-JUN-1999; 99US-0138598.

XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PS Claim 20; SEQ ID NO 15412; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA01910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 124 AA;
Query Match 1.4%; Score 7; DB 22; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 APEVPPP 75
DB 70 APEVPPP 76
RESULT 31
ID ABB55608 standard; Protein; 130 AA.
XX ABB55608;
XX 16-MAY-2002 (first entry)
DE Lactococcus lactis protein rpsI.
XX
KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
OS Lactococcus lactis IL1403.
XX
XX FR2807446-A1.
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolocrine A, Sorokine A, Renault P, Ehrlich SD;
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification or Lactococcus
PT lactis and related species -
XX
XX Claim 6; SEQ ID No 2310; 2504pp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the

CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 130 AA;
Query Match 1.4%; Score 7; DB 23; Length 130;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 447 GKTTVNG 453
DB 26 GKTTVNG 32
RESULT 32
ID AAG07374 standard; Protein; 144 AA.
XX AAG07374
XX AAG07374;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 4503.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135533.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142290.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145092.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148665.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.4%; Score 7; DB 21; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 477 TLN5FCN 483
DB 81 TLN5FCN 87

RESULT 33

AAC07373

ID AAC07373 standard; Protein; 149 AA.

XX AAC07373;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 4502.

XX Protein identification; signal transduction pathway; metabolic pathway;

KM hybridisation assay; genetic mapping; gene expression control; promoter;

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132487.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142820.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

```

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149920.
PR 25-AUG-1999; 99US-0149920.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match
Best Local Similarity 1.4%; Score 7; DB 21; Length 149;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 477 TLN5FGN 483
DB 86 TLN5FGN 92

```

RESULT 34
ID AAB25190
ID AAB25190 standard; Protein; 156 AA.
XX
AC AAB25190;
XX
DT 27-NOV-2000 (first entry)
XX
DE Eucalyptus grandis cell signalling involved protein SEQ ID NO:509.
XX
KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;
KW environmental change; development; cell proliferation; differentiation;
KW elongation; survival; disease resistance; nutrient metabolism.
XX
OS Eucalyptus grandis.
XX
PN MO200042171-A1.
XX
PD 20-JUL-2000.
XX
PF 11-JAN-2000; 2000MO-US00724.
XX
PR 12-JAN-1999; 99US-0228986.
PR 01-NOV-1999; 99US-0162866.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Strabala TJ, Nieuwenhuizen NJ;
XX
DR WPI; 2000-476052/41.
XX
PT Isolated polynucleotide encoding a polypeptide involved in cell
PT signaling used for generating transgenic plants with modified responses
PT to external signals.
XX
PS Claim 3; Page 225; 527pp; English.
XX
CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
CC pine (Pinus radiata also known as Monterey pine). The protein sequences
CC are involved in cell signalling. The polynucleotide and protein
CC sequences can be used to modify the response of plant cells to external
CC signals e.g. environmental changes or pathogens during the growth and
CC development of a plant. They can be used to modify cell proliferation,
CC differentiation, elongation and survival, resistance to disease and
CC nutrient metabolism. Examples of modifications which can be produced are
CC altered fruit ripening and senescence of leaves and flowers e.g. to
CC delay senescence and prolong the life of cut flowers or enhance
CC senescence of reproductive organs to engineer sterile plants. Other
CC modifications can be used to delay senescence in selected cell types or
CC organs providing fruit and vegetables which have a longer shelf life
CC between harvest and consumption, or to decrease branching frequency in
CC forest tree species giving long stretches of valuable knot-free clear
CC wood which can be used in solid timber furniture and veneers.
XX
SQ Sequence 156 AA;
XX
Query Match
Best Local Similarity 1.4%; Score 7; DB 21; Length 156;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 443 NGISGKI 449
DB 115 NGISGKI 121

```

RESULT 35
AAB58068
ID AAB58068 standard; Protein; 167 AA.
XX

AC	ABBS8068;
XX	
DT	26-MAR-2002 (first entry)
DE	Drosophila melanogaster polypeptide SEQ ID NO 996.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical.
OS	Drosophila melanogaster.
PN	WO200171042-A2.
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PP	11-JUL-2000; 2000US-0614150.
PA	(PEKE) PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	N-PESDB; ABL02171.
XX	
PS	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX	
CC	Discloure; SEQ ID NO 996; 21pp + Sequence Listing; English.
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01940-ABL16175) and the encoded proteins (ABBS7737-ABBS72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
SO	Sequence 167 AA;
XX	
QY	Query Match 1.4%; Score 7; DB 22; Length 167; Best Local Similarity 100.0%; Pred. No. 2.8e+02;
DB	Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0 72 VPPVDR 78 118 VPPVDR 124
RESULT 36	
ID	ABP38126 standard; Protein; 172 AA.
AC	ABP38126;
DT	24-JUL-2002 (first entry)
DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:2971.
XX	
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection; antibacterial; gene therapy.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	US6380370-B1.
XX	

30-APR-2002.
13-AUG-1998; 98US-0134001.
14-AUG-1997; 97US-055779P.
08-NOV-1997; 97US-064964P.
(GENO-) GENOME THERAPEUTICS CORP.
Doucette-Stamm LA, Bush D;
WPI: 2002-381255/41.
N-PSDB; ABN90671.
Novel isolated nucleic acid encoding a *Staphylococcus epidermidis* polypeptide, useful for diagnosing and treating bacterial infections -
Disclosure; SEQ ID 2971; 267bp; English.
ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP95124 to ABP37960. The *S. epidermidis* sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly *S. epidermidis* infections. The sequences can be used to screen for compounds able to interfere with the *S. epidermidis* life cycle or inhibit *S. epidermidis* infection.
N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site.
Sequence 172 AA;

Oy	219	PTTKGKY 225
Db	151	PTTKGKY 157
	RESULT 37	
	AA28428	
ID	AA28428	standard; Protein; 180 AA.
XX		
AC	AA28428;	
XX		
DT	15-FEB-2000	(first entry)
XX		
DE	Wheat branched chain amino acid transaminase amino acid sequence.	
XX		
KW	Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;	
KM	branched chain amino acid transferase; biosynthetic enzyme; antibody;	
KM	3-isopropylmalate dehydratase.	
XX		
OS	Triticum aestivum.	
XX		
Key	Location/Qualifiers	
FH	Key	
FT	Misc-difference 176	
FT	/note= "encoded by TAG"	
FT	Misc-difference 179	
FT	/note= "encoded by ATA"	
XX		
PN	W09921880-A2.	
XX		
PD	06-MAY-1999.	
XX		
PF	20-OCT-1998;	98WO-US22081.
XX		
PR	28-OCT-1997;	97US-0063423.
XX		
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.	

PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254037.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 DR N-PSDB; AAK59530.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metaabasis -
 XX
 XX Claim 11; SEQ ID NO 14342; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytotoxic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the

CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 180 AA;
 Query Match 1.4%; Score 7; DB 22; Length 180;
 Best Local Similarity 100.0%; Pred. No. 3e+02; Mismatches 0; Gaps 0;
 Matches 7; Conservative 0; Indels 0; Gaps 0;
 Qy 277 GPNLTSS 283
 Db 7 GPNLTSS 13
 RESULT 39
 ABB71959
 ID ABB71959 standard; Protein; 185 AA.
 AC ABB71959;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila, melanogaster polypeptide SEQ ID NO 42669.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL16062.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 42669; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL1840-ABL16175) and the encoded proteins
 CC (AAB57737-ABB72072).
 CC
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://ipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 185 AA;
 Query Match 1.4%; Score 7; DB 22; Length 185;
 Best Local Similarity 100.0%; Pred. No. 3e+02; Mismatches 0; Gaps 0;
 Matches 7; Conservative 0; Indels 0; Gaps 0;

QY 225 YGEOGLQ 231
 |||||
 Db 138 YGEOGLQ 144

RESULT 40

ABB71208
 ID ABB71208 standard; Protein; 187 AA.

AC ABB71208;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 40416.

KM Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

PI N-PSDB; ABL15311.

DR N-PSDB; ABL15311.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Disclosure; SEQ ID NO 40416; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL1840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 187 AA;

Query Match 1.4%; Score 7; DB 22; Length 187;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 LSGKITV 451
 |||||
 Db 56 LSGKITV 62

Search completed: August 27, 2003, 18:42:34
 Job time : 110 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 7, 2003, 00:14:41 / Search time 95 Seconds
(without alignments)
2332.361 Million cell updates/sec

Title: US-10-088-045-2
Perfect score: 2626
Sequence: 1 MSKPTLIKTKTLLICALSALML.....NKGQSLADVDVAKAKTKPN 502

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgml2_1/USPTO.spool/US10088045.rmlnat 04092003 083143 3945/app_query.fasta.1.647
-DB=Issued_Patents_NA -QPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALL=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTMT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10088045 @CCN 1.1 56 @unat 04092003 083143 3945 -NCPU=6 -ICPU3
-NO MAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGCLO
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	163.5	6.2	1938	4	US-09-252-991A-8682
2	163.5	6.2	1971	4	US-09-252-991A-8564
3	163	6.2	1329	3	US-09-252-991A-8564
4	163	6.2	1437	3	US-09-252-991A-8564
5	163	6.2	1437	3	US-09-252-991A-8564
6	158.5	6.0	4403765	3	US-09-103-840A-2
7	158.5	6.0	4411529	3	US-09-103-840A-1
8	155	5.9	1302	4	US-09-252-991A-6541
9	154.5	5.9	963	4	US-09-252-991A-6490
10	153	5.8	1470	4	US-09-252-991A-9169
11	153	5.8	1764	4	US-09-252-991A-8797
12	153	5.8	2256	4	US-09-252-991A-8910

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

13	152	5.8	1404	1	US-07-985-458-2	Sequence 2, Appli
14	151.5	5.8	1335	4	US-09-252-991A-765	Sequence 765, App
15	151.5	5.8	1524	4	US-09-252-991A-737	Sequence 737, App
16	151.5	5.8	1641	4	US-09-252-991A-646	Sequence 646, App
17	150	5.7	1554	4	US-09-252-991A-4893	Sequence 4893, App
18	150	5.7	2139	4	US-09-252-991A-4910	Sequence 4910, App
19	147.5	5.6	1977	4	US-09-252-991A-8339	Sequence 8339, App
20	146	5.6	1782	4	US-09-252-991A-14102	Sequence 14102, A
21	141	5.4	1588	2	US-08-706-037-24	Sequence 24, Appl
22	141	5.4	1588	2	US-09-005-397-124	Sequence 24, Appl
23	141	5.4	1672	1	US-08-172-331B-13	Sequence 13, Appl
24	136.5	5.2	980	4	US-09-615-192A-208	Sequence 208, App
25	132.5	5.0	1203	4	US-09-252-991A-8410	Sequence 8410, App
26	132	5.0	2016	4	US-09-328-352-1877	Sequence 1877, App
27	131.5	5.0	1908	4	US-09-328-352-3903	Sequence 3903, App
28	130.5	5.0	1323	4	US-09-252-991A-14169	Sequence 14169, A
29	130	5.0	3641	1	US-08-030-096-5	Sequence 5, Appli
30	130	5.0	6885	3	US-08-746-111-4	Sequence 4, Appli
31	128.5	4.9	6909	2	US-08-804-196-1	Sequence 1, Appli
32	128.5	4.9	6909	2	US-08-804-196-1	Sequence 1, Appli
33	128.5	4.9	6909	3	US-08-746-111-26	Sequence 26, Appl
34	128.5	4.9	6909	4	US-08-454-353A-1	Sequence 1, Appli
35	128.5	4.9	6909	4	US-09-165-019-1	Sequence 1, Appli
36	125.5	4.8	933	4	US-09-252-991A-6517	Sequence 6517, App
37	120.5	4.6	7032	3	US-09-324-867-1	Sequence 8062, App
38	119	4.5	1944	4	US-09-252-991A-8062	Sequence 8062, App
39	118.5	4.5	477	4	US-09-252-991A-6506	Sequence 1456, App
40	118.5	4.5	1494	4	US-09-252-991A-1456	Sequence 1456, App
41	118.5	4.5	1536	4	US-09-252-991A-1567	Sequence 1567, App
42	117.5	4.5	4665	4	US-09-297-937C-12	Sequence 12, Appl
43	117	4.5	3196	4	US-09-704-449-1	Sequence 5, Appli
44	114.5	4.4	1791	4	US-09-468-578-5	Sequence 1, Appli
45	114.5	4.4	1791	4	US-09-218-702-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-8682
; Sequence 8682 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: MARC J. RUBENFELD et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8682
; LENGTH: 1938
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8682

ALIGNMENT Scores:

Pred. No.: 1.99e-07 Length: 1938
Score: 163.50 Matches: 93
Percent Similarity: 33.96% Conservative: 52
Best Local Similarity: 21.78% Mismatches: 141
Query Match: 6.23% Indels: 141
DB: 4 Gaps: 21

US-10-088-045-2 (1-502) x US-09-252-991A-8682 (1-1938)

OY 19 MetLeuSerGlyCysSerSerGlnAlaAspIysAla---AlaGlnProIysSerSerThr 37
DB 121 GTTCTGCTGGGCTGTCACGCGCGGCTTCAGGCAATTGGCCGACAGGCTCCGACG 180

```

QY 38 ValaSPAlaAlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnIleGlnGly 57
DB -----CGCGGT 186
QY 58 GluLeuProValIle-----AspAlaIleValThrHisAlaProGluVal 72
DB -----CGCGGT 187
QY 187 GCCGCCCCCTGTGGTGAAGGCGACCGAGTTGCACTTGATCGCGCAATCGCCGCTG--- 243
QY 73 ProProProValaAspArgAspHisProAlaLysValValValLysMetGluThrValGlu 92
DB -----MACTTCAAGGCGACCGAGCGTGGCTAG--- 273
QY 93 LysValMetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlnGlyGlnVal 112
DB -----ACCATCAACGGCTCGCTG 291
QY 274 -----
QY 113 ProGlyGlnMetIleArgValArgGlnGlyAspThrIleGluValGlnPheSerAsnHis 132
DB -----CGCGCACGCAATTCGTTGGCGGAGGTGACCCGTCAACATCCGTGAACCAAC--- 348
QY 292 CGCGCACGCAATTCGTTGGCGGAGGTGACCCGTCAACATCCGTGAACCAAC--- 348
QY 133 ProAspSerLysMetProHisAsnValAspPheHisAlaAlaThrGlyPro----- 149
DB -----
QY 349 ---AAGTGGGTGAAGCCACATTCATTCATCGACGAGATCATCTCGGTATCAGATG 405
QY 150 GlyGlyGlyAlaGlnAlaAspPheThr-----AlaProGlyHisThrSerThrPheSer 167
DB -----GATGGCGGTGGCGGCGCATCAGCTTCATGCGCATCGCTCCGCGGACGATTCAGTACCGC 465
QY 406 GATGGCGGTGGCGGCGCATCAGCTTCATGCGCATCGCTCCGCGGACGATTCAGTACCGC 465
QY 168 PheLysValaLeuGlnProGlyLysThrValTyrHisCysAlaValaProValGlyMet 187
DB -----TTCAAGGTTGACGACGAGCGGTCTCTACTGTGACAC-----TCGCACTCGCGGTTC 516
QY 466 TTCAAGGTTGACGACGAGCGGTCTCTACTGTGACAC-----TCGCACTCGCGGTTC 516
QY 188 HisIleAlaAsnGlyMetTyrGlyLysLeuValGluProLysGlnGlyLysProLys 207
DB -----
QY 517 CAGGAATCTCAGCGGCAATGATCCAAAGCGACATCATCATCAACCC---GCTGGCGCGAGAC 573
QY 208 Val-----AspLysGluTyrTyrValMet----- 215
DB -----
QY 574 ATCAGCGCGCGATGACGACGAGTGTCTGTTCCGATTGAGACGACGAAAGCCGATG 633
QY 216 -----GlnGlyAspPheTyrThr----- 221
DB -----
QY 634 CGGGTGTTCACCAAGTCAAGTCCAAAGCGACATTCATTCACCAACCAACCTACCGTC 693
QY 222 -----LysGly 223
DB -----
QY 694 TTGCATTTCTCCGGGACGCGCTCGCGGAAAGCACGCGCTCGATGACAAAGCGGACG 753
QY 224 LysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla 243
DB -----
QY 754 ATGTGAAACGAGATGCGGATGAAATCCGACGATCTGTGCGACACGCTG 813
QY 244 GlnTyrValValaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAla 263
DB -----
QY 814 ACCTATCTGGCC---AAGCGGTGACCCCGCC---GSCAATTGAAAGCGCGCTGTC 864
QY 264 LysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSer 283
DB -----
QY 865 CGACCGGTGACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 924
QY 284 PheHisValIleGlyGluIlePheAspLysValaHisPheGlnGlyGlyLysGluAsn 303
DB -----
QY 925 GTACGATTCGACGATGAAAGTCAAGTCAAGTGTGCGATGCTGCTGCTGCTGCTGCTG 969
QY 304 HisAsnIleGlnThrThrLeuLeuProAlaGlyAlaAlaIleThrGluPheLysVal 323
DB -----GTT 972
QY 970 -----
QY 324 AspValProGlyAspTyrValLysValaAspHisAlaIlePheArgAlaPheAsnLysGly 343
DB -----
QY 973 GACGTG---GAGCGGTGACCGGTGACG---TTCCGA---TTCCGCTCGCTGCTGCTG 1017
QY 344 ---AlaLeuGlyLysLeuLysValaGluGlyGluAsnHisGluIleTyrSerHisLys 362

```

```

DB 1018 GAAACCGTCGACGTGATCTGCTCCAGCCAGGATGACGCTTACAGATCTTCCGCCAGGCG 1077
QY 363 GlnThrAspAlaValaLysLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAla 382
DB 1078 ATGACCGGACAGGCTAC-----GCCGTGCCACGCTGCGCTGCGTGAAGGT 1125
QY 383 ProLysThrProAlaProAla 389
DB 1126 CTTCAGCTCCCGCTGCTGCC 1146

RESULT 2
US-09-252-991A-8564/c
; Sequence 8564, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8564
; LENGTH: 1971
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8564

Alignment Scores:
Pred. No.: 2,04e-07 Length: 1971
Score: 163.50 Matches: 93
Percent Similarity: 33.96% Conservative: 52
Best Local Similarity: 21.78% Mismatches: 141
Query Match: 6.23% Indels: 141
DB: 4 Gaps: 21

US-10-088-045-2 (1-502) x US-09-252-991A-8564 (1-1971)
QY 19 MetLeuSerGlyCysSerAsnGlnAlaAspLysAla---AlaGlnProLysSerSerThr 37
DB -----
QY 1956 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1897
QY 38 ValaSPAlaAlaAlaLysThrAlaAsnAlaAspAsnAlaAlaSerGlnGlnIleGlnGly 57
DB -----CGCGGT 1891
QY 1896 -----
QY 58 GluLeuProValIle-----AspAlaIleValThrHisAlaProGluVal 72
DB -----
QY 1890 GCCGCCCCCTGTGGTGAAGGCGACCGAGTTGCACTTGATCGCGCAATCGCCGCTG--- 1834
QY 73 ProProProValaAspArgAspHisProAlaLysValValValLysMetGluThrValGlu 92
DB -----MACTTCAAGGCGACCGAGCGTGGCTAG--- 1804
QY 1833 -----AAGTTCACGCGGACCGAGCGTGGCTAG--- 1804
QY 93 LysValMetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlnGlyGlnVal 112
DB -----ACCATCAACGGCTCGCTG 1786
QY 1803 -----
QY 113 ProGlyGlnMetIleArgValArgGlnGlyAspThrIleGluValGlnPheSerAsnHis 132
DB -----CGCGCACGCAATTCGTTGGCGGAGGTGACCCGTCAACATCCGTGAACCAAC--- 1729
QY 1785 CGCGCACGCAATTCGTTGGCGGAGGTGACCCGTCAACATCCGTGAACCAAC--- 1729
QY 133 ProAspSerLysMetProHisAsnValAspPheHisAlaAlaThrGlyPro----- 149
DB -----
QY 1728 ---AAGTGGGTGAAGCCACATTCATTCATCGACGAGATCATCTCGGTATCAGATG 1672
QY 150 GlyGlyGlyAlaGlnAlaAspPheThr-----AlaProGlyHisThrSerThrPheSer 167

```

```

Db      1671 CATGCGTTCGGGCGCATCAGCTTCAATGCGATCGCTCCGGCGAGACCTTCAAGTACCGC 1612
Qy      168 PheLysAlaLeuGlnProGlyLeuTyValTyrHisCysAlaValAlaProValGlyMet 187
      |||||
Db      1611 TTCAGGTTCCAGACAGCGGCTCTACTGTTACAC-----TCGCACTCCGGCTTC 1561
Qy      188 HisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlnProLysGlnGlyLeuProLys 207
      |||||
Db      1560 CAGCAATCAGCGGCGCATGTATCGGGCGCATCATCATCAGACCC--GCTGGCGCGAGACC 1504
Qy      208 Val-----AspLysGlnTyrTyrValMet----- 215
      |||||
Db      1503 ATTCACCGCGCGATGACCAAGTCTGCTGTTCTCGGANTTGCACGACGAGACCCGATG 1444
Qy      216 -----GlnGlyAspPheTyrThr----- 221
      |||||
Db      1443 CGGGTTTACCAAGTCAAGTCCCAAGGAGCTATTACACTACACCAACCTACCGTC 1384
Qy      222 -----LysGly 223
      |||||
Db      1383 TTGCACTTCTCCGGAGCCCTTCGCGAAGCGACCGCGCTGCGATCGACCAAGCGCAAG 1324
Qy      224 LysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGlnLysAlaIleArgGlnAspAla 243
      |||||
Db      1323 ATGTGACACGATGCGGATGAATCCGACGATCTGCGAGATCTGTCGCCACACGCTG 1264
Qy      244 GlnTyrValValPheAsnGlySerValGlyAlaLeuThrGlyGlnAsnAlaLeuLysAla 263
      |||||
Db      1263 ACCTATCTGCCC--AAGCGCGTCAACCCCGCC--GGCAATTGACGCGCGCTTC 1213
Qy      264 LysValGlyGlnThrValArgLeuPheValGlyAsnGlyLysProAsnLeuThrSerSer 283
      |||||
Db      1212 CACCGCGGTGACGCGCTGCGCGCTGCGCATGTCACGCGTGCAGCAACCTTCTACGAC 1153
Qy      284 PheHisValIleGlyGlnLeuPheAspLysValHisPheGlnGlyGlyLysGlnLys 303
      |||||
Db      1152 GTACGCAATCCAGGACTGAACCTCAGCGGTGTCACGTGCTG----- 1108
Qy      304 HisAsnIleGlnThrThrLeuIleProIleGlyAlaIleAlaIleThrGlnPheLysVal 323
      |||||
Db      1107 -----GTT 1105
Qy      324 AspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGly 343
      |||||
Db      1104 GACGTG-----GACCGGTGACCGTGCATGAG-----TTCGA--TTCGCTCTGCG 1060
Qy      344 ---AlaLeuGlyIleLeuLysValGlnGlyGlnAsnHisGlnLeuTyrSerHisLys 362
      |||||
Db      1059 GAACCGTGCAGTGTATCTCCAGCCACCGGATGAGCTTACGATCTTCCGCCCGAGCG 1000
Qy      363 GlnThrAspAlaValTyrLeuProGlnGlyAlaProGlnAlaIleAspThrGlnGlnAla 382
      |||||
Db      999 ATGACCGGACAGGCTAC-----GCGGTGCCACGCTGCGCGCTGAGGT 952
Qy      383 ProLysThrProAlaProAla 389
      |||||
Db      951 CTTCAAGCTCCCGTCTGCTCC 931

```

```

; SEQ ID NO 23
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-23

Alignment Scores:
Pred. No.: 1,25e-07 Length: 1329
Score: 163.00 Matches: 118
Percent Similarity: 35.85% Conservative: 34
Best Local Similarity: 27.83% Mismatches: 144
Query Match: 6.21% Indels: 120
DB: 3 Gaps: 25

US-10-088-045-2 (1-502) x US-09-296-284-23 (1-1329)

Qy      171 LeuGlnProGlyLeuTyrValTyrHis-----CysAlaValAlaPro 184
      |||||
Db      22 ATTCAGCGCGGTGCTTACGTCGCGCGCTGCTGACTGCGTTCGATACCGACTA 81
Qy      185 ValGlyMetHisIleAlaAsnGlyMet-----TyrGly 195
      |||||
Db      82 CACGCGCACCTTTGCTGCTGCTGCTGATCAGAGCCGATCGCGACGATCTACTCC 141
Qy      196 Leu---IleLeuValGlnProLysGlnGlyLeuProLysValAspLysGlnTyrVal 214
      |||||
Db      142 ACCAATATCAGCGCTGACCCGAAATACGGTATCGGCAC--TATACA 186
Qy      215 MetGlnGlyAspPheTyrThrLysGlnLysTyrGlyGlnGlnGlyLeuGlnProPheAsp 234
      |||||
Db      187 CTCGAAAGAT--TTCAGAAAGCG----- 207
Qy      235 MetGlnLysAlaIleArgGlnAsp----- 242
      |||||
Db      208 ATCCGTAAGGATATCCGCAAGACGCGCGGATTTACCGGCATGCGCTATCTGAG 267
Qy      243 -----AlaGlnTyrValValPheAsnGlySerVal 252
      |||||
Db      268 TTCGCTGCGCTGATGATGACGATCAAGGCCATGTATGCTTTCATGATGCGCTG 327
Qy      253 GlyAla-----LeuThrGlyGlnAsnAlaLeu 261
      |||||
Db      328 AA-GCGCGTCCGCTTCAAGAACAGCGCGGACATCTTCCGCGCATMAACATGCGCTG 386
Qy      262 LysAlaLysValGlyGlnThrValArg--LeuPheValGlyAsnGlyLysProAsnLeu 281
      |||||
Db      387 -----GCCGTTGGCATCTGCGCGCGCATGTTGTTCCGACTGTACACCGGCTTCA 440
Qy      281 rSerSerPheHis-----ValIleGlyGlnLeuPheAspLysValHisPheGln 297
      |||||
Db      441 CAAGAGCATCTCCGATCCGAGATGCGCGCTGCGCAATACCTC-----GTAAATGCGCC 494
Qy      297 uGlyGlyLysGlyGlnAsnHisAsnIleGlnThrLeuIleProAlaGlyAlaAla 317
      |||||
Db      495 AGCGCATCTGTGCGAGTGCATACG-----CCCGGTGCATGCGCAT 536
Qy      317 AlleThrGlnPheLysValAspValProGlyAspTyrValLeu-----Val 332
      |||||
Db      537 GCAAGTCAAGGCTTATACGCGCCAGAGACGCAACGCTTACCTTCGCGTCCGACCGAT 596
Qy      332 AspHisAlaIle-----PheArgAlaPheAsnLysGlyAlaLeuGly----- 346
      |||||
Db      597 CGACAACTGATGCTCCAGCTCCGATGACAAATAGCGACAGCGGCTTGGTGGCTGTC 656
Qy      347 -----IleLeuLysValGlnGlyGlnAsnHisGlnLeuTyrSer 360
      |||||
Db      657 TGAAGCGACATGCGCGATTCCTGAG--AGCGGCTTATCGACCATTCGCGCTTC 713
Qy      360 rHisLysGlnThrAspAlaVal---Tyr--LeuProGlnGlyAlaProGlnAlaIleAsp 379
      |||||
Db      714 CGGTGCAATGCTGACGTGTGTGCTTACGACCCAGCACTGACGACGACGATTTGCA 773
Qy      379 hrGlnGlnAla-----ProLysThrProAlaProAlaAsnLeu--G 392

```

```

Db      774 CGCAACG-GCCAACTGTAAGAGCATGCGCGCTTCGGAAGCAAAAACCTGGGTC 832
Qy      392 InGIuGInIleYs-----AlaGIlyYsA 400
Db      833 AGGATGACGGCAAGCCACGCGCTGCTCGAAGCCGGTGCAAGGATGATGACGGCCAG 892
Qy      400 lATHrTYrAsPSeRAsnCYSAlaAlaCYshISgInProAsPGLySGlyValProAsnA 420
Db      893 AGGTTTACCTCCAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 952
Qy      420 lApHeProPLeuAlaAsnSerAsPtyrLeuAsnAlaAsPHisAlaArgAla--AlAs 439
Db      953 TGTTCGGCGCGCTGGTGGCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1012
Qy      439 eRIleValAlaAsnGlyLeuSerGlyYsIleThrValAsnGlyAsnGIntYrGlySerV 459
Db      1013 ACATCGGACATTCGGC-----GGTATTCTGCTCCGACGAATACGGCCGATCTGCTG 1066
Qy      459 aL---MePProAlaIleAla-----LeuSerAsPGLInIleAlaAsnValIleThrT 476
Db      1067 TTGCCATGCGCGGCTTCGCGCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1126
Qy      476 yThrIleuAsnSerPheGlyAsnLySGly---GlyGlnLeuSerAlaAsPAspValAla 495
Db      1127 TCATGCCCAAGAGCTGGGGCAACGAGGCTCGGGAACCTGCTGCTGCTGCTGCTGCTGCA 1186
Qy      495 ySaIaLyS 497
Db      1187 AGCTCCGC 1194

```

RESULT 4

US-09-296-284-2

Sequence 2, Application US/09296284A

Patent No. 6204040

GENERAL INFORMATION:

APPLICANT: Choi, Eun-Sung

APPLICANT: Rhee, Sang-Ki

APPLICANT: Lee, Eun-Hae

TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes

FILE REFERENCE: 1533.0870000

CURRENT APPLICATION NUMBER: US/09/296,284A

CURRENT FILING DATE: 1999-04-22

NUMBER OF SEQ ID NOS: 87

SOFTWARE: Patentn Ver. 2.0

SEQ ID NO 2

LENGTH: 1437

TYPE: DNA

ORGANISM: Gluconobacter suboxydans

US-09-296-284-2

Alignment Scores:

```

Pred. No.: 1,41e-07 Length: 1437
Score: 163.00 Matches: 118
Percent Similarity: 35.85% Conservative: 34
Best Local Similarity: 27.83% Mismatches: 144
Query Match: 6.21% Gaps: 130
DB: 3 Gaps: 25

```

US-10-088-045-2 (1-502) x US-09-296-284-2 (1-1437)

```

Qy      171 LeuGInProGlyLeuTYrValTYrHis-----CysAlaValAlaPro 184
Db      130 ATTAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 189
Qy      185 ValGlyMeHisIleAlaAsnGlyMe-----TYrGly 195
Db      190 CAGGCGCAGCTTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
Qy      196 Leu---lIleuValGluProLySGlyGlyLeuProLySValAsPAspValAla 214
Db      250 ACCAACATCAAGCTGACCGCAATACGGTATCGGCAAC-----TATACA 294

```

```

Qy      215 MetGInGlyAsPheThrTYrThrLySGlyTYrGlyGlyGlnGlyLeuGInProPheAsP 234
Db      295 CTCGAAGAT-----TTCAGAAAGCG----- 315
Qy      235 MetGlyLySAlaIleArgGluAsP----- 242
Db      316 ATCCGTAAAGGATTCGCAAGAGCAGGCGGACGGTTTATTCGCGCATGCCGTATCTGAG 375
Qy      243 -----AlaGIlyTYrValAlaPheAsnGlySerVal 252
Db      376 TTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 435
Qy      253 GlyAla-----LeuThrGlyGlnAlaLeu 261
Db      436 AA-GCCGCTGCGCTTCAGAACAGACCGGACATCTCTGCGCCATGAACATGCGCTG 494
Qy      262 LySAlaLySValGlyIleThrValArg-LeuPheValGlyAsnGlyLyProAsnLeuTh 281
Db      495 -----GCCGTTGGCATCTGCGCGCGATGTTGTTCCGACTGTCAACACGAGCTCGA 548
Qy      281 rSerSerPheHis-----ValIleGlyGlnIlePheAsPlyValHisPheG 297
Db      549 CAAGACCATCTCGATCCGGAAGTGGCGGCGGCAATACCTC-----GTGAATGAGCC 602
Qy      297 uGIlyGlyLySGlyAsnHisAsnIleGlnThrThrLeuIleProAlaGlyValAla 317
Db      603 AGGCCATTGTGGCGAGTGTATTCG-----CCCCGTGGCATGGCCAT 644
Qy      317 aIleThrGluPheLySValAsPValProGlyAsPtyrValLeu-----Va 332
Db      645 GCAGTCAAGGGGTATACGGCCAAAGACGGCAACGTTTACTCTCCGTGGCGCACCGAT 704
Qy      332 lAsPHisAlaIle-----PheArgAlaPheAsnLySGlyAlaLeuGly----- 346
Db      705 CGCAACTGATGATCTCCAGCGCTGGTGCATACGACGACGCGTGGTGGCTGGCTGCTG 764
Qy      347 -----lIleuLySValGlyGlyGlnGlyAsnHisGlyIleTyrs 360
Db      765 TGAAGACGACATTGCCGAGTCTCTGAAG--AGCGCCGATTCGACCATTCGCGCTT 821
Qy      360 rHisLySGlnThrAsPAlaVal---TYr-LeuProGlnGlyAlaProGlnAlaIleAsP 379
Db      822 CGGTGCATGCGTGAAGTGGTGGCTTACAGACCCAGCATGGACGACGACGATCTGCA 881
Qy      379 hrcGlnGluAla-----ProLySthrProAlaPAspValLeu---G 392
Db      882 CGCAACG-GCCAACTACTGAAGAGCATGCGCGCTTCGGAAGCAAAAACCTGAGTC 940
Qy      392 InGIuGInIleYs-----AlaGIlyYsA 400
Db      941 AGGATACGGCAAGGCCACGCGCTGCTCGAAGCCGCTGCGCAAGGTGATGCGGCGCAG 1000
Qy      400 lATHrTYrAsPSeRAsnCYSAlaAlaCYshISgInProAsPGLySGlyValProAsnA 420
Db      1001 AGGTTTACCTCCAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1060
Qy      420 lApHeProPLeuAlaAsnSerAsPtyrLeuAsnAlaAsPHisAlaArgAla--AlAs 439
Db      1061 TGTTCGGCGCGCTGGCTGGCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1120
Qy      439 eRIleValAlaAsnGlyLeuSerGlyYsIleThrValAsnGlyAsnGIntYrGlySerV 459
Db      1121 ACATCGGACATTCGGC-----GGTATTCTGCTCCGACGAATACGGCCGCTGCTGCTG 1174
Qy      459 aL---MePProAlaIleAla-----LeuSerAsPGLInIleAlaAsnValIleThrT 476
Db      1175 TTGCCATGCGCGGCTTCGCGCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1234
Qy      476 yThrIleuAsnSerPheGlyAsnLySGly---GlyGlnLeuSerAlaAsPAspValAla 495
Db      1235 TCATGCCCAAGAGCTGGGGCAACGAGCTCCGGAACCTGCTGCTGCTGCTGCTGCTGCTGCA 1294

```


? APPLICANT: FLEISCHMAN, Robert D.
 ? APPLICANT: WHITE, Owen R.
 ? APPLICANT: FRASER, Claire M.
 ? APPLICANT: VENTER, John C.
 ? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ? TITLE OF INVENTION: TUBERCULOSIS
 ? FILE REFERENCE: 24366-2007.00
 ? CURRENT APPLICATION NUMBER: US/09/103,840A
 ? CURRENT FILING DATE: 1998-06-24
 ? NUMBER OF SEQ ID NOS: 2
 ? SOFTWARE: PatentIn Ver. 2.1
 ? SEQ ID NO 1
 ? LENGTH: 4411529
 ? TYPE: DNA
 ? ORGANISM: Mycobacterium tuberculosis
 ? OTHER INFORMATION: H37Rv
 ? US-09-103-840A-1

Alignment Scores:
 Pred. No.: 0.12 Length: 4411529
 Score: 158.50 Matches: 104
 Percent Similarity: 31.00% Conservative: 51
 Best Local Similarity: 20.80% Mismatches: 180
 Query Match: 6.04% Indels: 165
 DB: 3 Gaps: 23

US-10-088-045-2 (1-502) x US-09-103-840A-1 (1-4411529)
 Oy 16 SerAlaLeuMetLeuSerGlyCySeSerAenGlnAlaAspLysAlaIaGlnProlySer 35
 Db 944111 AGCGGCTTGCGCTGAGCGGCTTGC-----CCCTCGAAGCCACGCGCA 944070
 Oy 36 SerThrValAspAlaIaIaLysThrAlaAsnAlaAspAlaA--AlaSerGln 54
 Db 944069 TCC-----GGGCGCGCGCGGATGACCGCTGCAATGACGCGCGCGAGCGCGCGCGCG 944016
 Oy 55 HisGlnGlyGlnLeuProValIleAspAlaIleValThrHisAlaProGlnValProPro 74
 Db 944015 CACAGTGGCGCA-----ACGGTTACCGCCACGCTAC-----CCCCAG 943977
 Oy 75 ProValAspArgAspHisProAlaLysValValValMetGlnThrValGlnLysVal 94
 Db 943976 CCGGCGAGATGACCTGGTGGCGCGATCTGCACACCTG----- 943935
 Oy 95 MetArgLeuAlaAspGlyValGluTyrGlnPheThrPheGlyGlyGlnValProGly 114
 Db 943934 -----ACCTACGGCAACACCATCCCGGA 943911
 Oy 115 GlnMetIleArgValArgGlnGlyAspThrIleGlnValGlnPheSerAsnHisProAsp 134
 Db 943910 CCACGATCCCGGCCACCCCTCGGGATGAGATTGTGCTCGGTGACCAACCTCTGGGT 943851
 Oy 135 SerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyProGlyGly 152
 Db 943850 GAT-----CCGACGTGGGTGCAATTGGCAGCGGATCGCGTGGCGCAACGATATGATGGC 943797
 Oy 153 AlaGlnAlaSerPheThr-----AlaProGlnHisThrSerThrPheSerPheLysAla 170
 Db 943796 ACCGAGCCCGCGACATGCGCAACATCGGCCCGCGGCGTGAAGATTACGATACCGGTTCTCCG 943737
 Oy 171 LeuGlnProGlyLeuTyrValTyrHisCysValaValAlaProValaIaIaMetHisIleAla 190
 Db 943736 CCGGATCCCGGACCTACTACGGGCCCTCGGACATCGCATGCTGCTTCMAAGCGCAC----- 943683
 Oy 191 AenGlyMetTyrGlyLeuIleLeuValGlnProLysGlnGlyLeuProLysValaAspLys 210
 Db 943682 ---GGCTTATCTGCGCTTCTGCTGTCGACGATCGCACTGAGCCAGGCGCACTACGACGCC 943626
 Oy 211 GlnTyrTyrValMetGlnLysAspPheTyrThrLysGly----- 223
 Db 943625 GAATGATCATCATCTCTGACGAT---TGGAGCGAAGGATCGGGAAGTCCCGGCAACAG 943569
 Oy 224 LysTyrGlyGlnGlnGlyLeuGlnProPheAspMetGlnLysAlaIleArgIuaAspAla 243

```

Db      943568 CTCTACGGCGAAGCTG-----ACCGACCSCCAAAACCCACCATATGCATAAACACA   943521
OY      243 -----|||||-----|||-----|||-----|||||-----
Db      943520 ACAGGTATGCCCGGAAGCGGTTGACAGCACACTGCTGGCGCGACGAGGAGGAC   943461
OY      244 -----GATTGTATATATPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla   260
Db      943460 ATCGCCTTACCGGTAATCTTGATTCMAAGGGCGAATCCCC-----GTGGCGCCACCGTCT   943407
OY      261 LeuValAlaIleValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu   280
Db      943406 TTTTAGGCGCAAGCTGGCGCACGCAATCGGATCGGATCGCATCTCAAACAGCGCCGCGACACC   943347
OY      281 ThrSerSerPheHisValIleGlyIlePheAspLysValHisPheGlyGly-----298
Db      943346 GCGTTCCGCATCGCGCTGCGCGGGCATTCGATGACGGTCAACCCACACCGACGGTTACCCA   943287
OY      299 -----GTLyVgLygluAsnHisAsnIleGln   307
Db      943286 GTGATTTCCACCGAAGTGCACGCTGTGATGCTGCGTGGCCCAACCGCTCGACGTCATG   943227
OY      308 ThrThrLeuLeuProAlaGlyGly-----AlaAlaIleThrGluPheLysVal   323
Db      943226 GTGACCC-----GCGCGTGGCGGGGTCTTTCCCTCGTGCACACTCGCGGAGAAGCGAAC---   943176
OY      324 AspValProGlyAspTrpValLeuValAspHisAlaIlePheArgAlaPheAsnLysGly   343
Db      943175 -----AAGCGGTGGCGGGTGGCGTGGCTGTCTACCC   943146
OY      344 AlaLeuGlyIleLeuLysValGlyGlyGluAsnHisGluIleTrpSerHisLysGln   363
Db      943145 GCGCGCCCGCG-----AGCCCA   943131
OY      364 ThrAspAlaValTrpLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaPro   383
Db      943130 CCCGACCCCGGAGTTTGGCGCGGAGTGAATCTCACTGGCGAGATGGGTACCTGTGAAATGTTTC   943071
OY      384 LysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLysAlaThrTrpAsp   403
Db      943070 ACGCGCCGACAGCATGCGCACTCG-----943047
OY      404 SerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAlaPhePro   423
Db      943046 -----GGCGGCGCCGAACCCACCGACGACTCCGC   943017
OY      424 LeuAlaAsnSerAspTrpLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsn   443
Db      943016 -----GTCACCTTGGGC   943005
OY      444 GlyLeuSerGlyLysIle-----ThrValAsnGlyAsnGlnTrpGluSerValMetPro   461
Db      943004 GGCAACCATGCGCAATGTAACGACTGACATCAACGGGGAAACCTTACACGACCAATCCA   942945

RESULT 8
US-09-252-991A-6541/c
Sequence 6541, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6541
LENGTH: 1302
```



```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6541

Alignment Scores:
Pred. No.:      8,46e-07      Length:      1302
Score:          155.00      Matches:      87
Percent Similarity: 37.43%      Conservative: 38
Best Local Similarity: 26.05%      Mismatches: 109
Query Match:      5.90%      Indels:      100
DB:              4          Gaps:      16

US-10-088-045-2 (1-502) x US-09-252-991A-6541 (1-1302)

QY 115 GlmetlleargvalagsglulglYAspThrile-----gluValglnPheSerAsnHis 132
DB 1216 AAGCTTTAGAGAGTAAAGAAATGACGACCTTTGAGTCTGTAACTTACAGCTCTCAC 1157
QY 133 ProAspSerLysMetProHisAsnValAspPheHisAlaIathrglYProglYglY 152
DB 1156 CCT-----CGGACCTGCTGGCGGCGCTGACCTGCTATCTTCCGCAACCGCAAGGCCA 1103
QY 153 Alaglu--AlaserPheThrAlaProglYhiSTrSerThrPheSerPheLysAlaleu- 171
DB 1102 GCGCAGCAGACACACCGACGACGCGGCGCCTTACGACGCGCATCGAGAGTACGA 1043
QY 172 GlNProglYleuTYrValTYrHisCyAlaValAlaProValglYmethis-----Ile 189
DB 1042 CAACCCCTGCTCCGAAGTG-----GTGGTTATGCTGTTCTCGGCAACCTGCTGTT 992
QY 190 AlaAsnglYmetTYrglYleuile-----leuValgluProLysglulY 204
DB 991 CGCGGTGGGCTACTGCGCCCTGTACCCGGCGCTGGGACCTGGAAAGGCTGATGCCGG 932
QY 205 leuPro-LysValAlaAspLysglulTYrTYrValmetGlnLYAspPheTYrThrLysglY 224
DB 931 CTACCACTCCGCCAC-----GAATTCGCCGCAAGGAAAA 896
QY 224 sTYrglYglulglYleuGlnProPheAspMetGluLYAlaIleargLysAlaI 244
DB 895 A---GGCTGACCGGCTCCACCACTGAGGAAAGAAATGACC---AAGGCCAGCAGAA 842
QY 244 uTYrValValAlaPheAsnLYSerValglYAlaIleuThrglYgluAsnAlaleuYs---- 262
DB 841 ATACGCCGCCGATCTTCCCAAGTTCGCCGCGATGCCATCGAGAAATCGCAAGATCC 782
QY 263 -----AlaLYValAlaIglulThrValAlaGluPheValglYAsn----- 275
DB 781 GCAGGCGGTGAATAATGGCGGT-----CGCTGTGCTGCTCCCAACTGCTGATCTGCA 728
QY 276 -----glYglYProAsnleuThrSerSerPheHisVal11 287
DB 727 CGGCTCGACGACCAAGGCGGCTAGCGCTTCCCACTGACCAACCGCACTGGCGCTG 668
QY 287 eglYglullePheAspLysValHisPheglulglYLYSerLYgluAsnHisAsn1leg1 307
DB 667 GGGC-----GGCGAGCGGAAACCAATCA 644
QY 307 nThrThrleuileProAlaIglYAlaAlaIleThrGluPheLysValAlaProgl 327
DB 643 GACCAACCATCATG--GCTGGC----- 625
QY 327 yAspTYrValleuValAspHisAlaIlePheargAlaPheAsnLYglYAlaleuGlY1 347
DB 624 -----CGCCAGCGCCGCAATGCCGCTGGGCTGA-----GT 593
QY 347 eleuLYsValAlaIglulgluAsnHisglulYSerHisLYSerHisLYThrAspAlaVala 367
DB 592 GATCGCGAGAGAGAGCGCGTGAAGAAACGTGCGGCTTGTGCTCTCCCAAGATGATGGCG 533
QY 367 lTYrleuProglulglYAlaProglAlaIleAspThrGlnGlulAlaProLysThrProAl 387
DB 532 TAAGCTGCGGAGGCGCC----- 514
```

```

QY 387 aProAlaAsnleuGlnIglulglYleuLYsValAlaIleThrTYrAspSerAsnCYsAl 407
DB 513 -----AAGCAGACATTCAGGCCCGGCAAGCAGCTTGTGCTGACCTGCGT 467
QY 407 aAlaCYsHisgluProAspLYLysglYValPro 418
DB 466 CGCCTGCCACGCTCCGAAAGGCAAGGCAAGCCCG 433

RESULT 9
US-09-252-991A-6490
; Sequence 6490, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6490
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6490

Alignment Scores:
Pred. No.:      5,96e-07      Length:      963
Score:          154.50      Matches:      87
Percent Similarity: 37.20%      Conservative: 35
Best Local Similarity: 26.52%      Mismatches: 107
Query Match:      5.88%      Indels:      99
DB:              4          Gaps:      16

US-10-088-045-2 (1-502) x US-09-252-991A-6490 (1-963)

QY 120 ArgglulYAspThrileglu--ValglnPheSerAsnHisPheAspSerLysMetPro 138
DB 3 AAGATGAGACACTTGTGAGTCTGTAACTTACAGCTCTCACCT-----CGGCAACCT 56
QY 139 HisAsnValAspPheHisAlaIathrglYProglYglYAlaIleu--AlaserPheT 158
DB 57 GCTGGCGCTGACCTGCTGATCTTCCGCAACCGCAAGGCCAGCAGACACCGA 116
QY 158 hTYrAlaProglYhiSTrSerThrPheSerPheLysAlaleu-GlnProglYleuTYrVal 177
DB 117 CGAGACCGTGGCGCATCTCTACGACGCGCATCGAGAGTACGACAACTCCCTGCCAAGTG 176
QY 178 TYrHisCyAlaValAlaProValAlaGluTYrMetHis-----IleAlaAsnglYmetTYrglY 195
DB 177 -----GTGGTTATGCTGTTCTGCGGACCTGAGTTCGCGGTGCGCTGACTGCTGCG 227
QY 196 leuile-----leuValgluProLysglulglYleuPro-LysValAspLY 210
DB 228 CTTGATCCCGGCGCTGGGCACTGGAAGAGCTGATGCCGGGTATCACTGCCGCAAC-- 285
QY 210 sgluTYrTYrValmetGlnLYAspPheTYrThrLysglYLYSerLYgluGlnLYle 230
DB 286 -----GAATTCGCCGCAAGGAAAAA---GCTGAGACCGGCT 320
QY 230 uGlnProPheAspMetGluLYsAlaIleargLysAlaIlegluTYrValAlaPheAsn1 250
DB 321 CCACAGTGGGAAAAAGAAATGAGCC---AAGGCCGACGAAATATACGCCCGCATCTTCCG 377
QY 250 ySerValglYAlaIleuThrglYgluAsnAlaleuYs-----AlaLYsValgl 266
DB 378 CAAGTTGCGCGCATGCTCCATCGAGAAAGTCCGCAAGGATCCGCAAGCGGTGAATAATGG 437
```

QY 266 ygluThrValArgLeuPheValGlyAsn----- 275
DB 438 CGGT-----CGCTGTTCCCTCTCACTGCAATGCGCAGCGCTCCGACCGCAAGG 491
QY 276 -----GlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluLeuPheAsp 293
DB 492 CGCTTACGGGTTTCCCACTGACCGACGCGCATGCGCGCTGGGGC----- 537
QY 293 sValHisPheGluGlyGlyLeuSerHisAsnIleGlnThrThrLeuLeuProAl 313
DB 538 -----GCGACGCGCGAAACCATTAACACCATCTCATG---GC 572
QY 313 aglGlyAlaAlaIleThrGluPheValAspValProGlyAspTyrValLeuValAs 333
DB 573 TGGC-----CG 578
QY 333 pHisAlaIlePheArgAlaPheAsnIleGlyAlaLeuGlyIleLeuLysValGluGly 353
DB 579 CCAAGCGCGCGCATGCGCGCTGGGTGAA-----GTGATCGCGCAGGAGAGCGCT 626
QY 353 uGluAsnHisGluIleTyrSerHisIleGlnThrAspAlaValTyrLeuProGluGlyAl 373
DB 627 GAAGAACGTCGCGCGCTTCTGCTCTCACCGACGATGCGCGCTGACGCTGCGAAGCGC 666
QY 373 aProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGln 393
DB 687 C-----AAGC 692
QY 393 uGlnIleLysAlaGlyLysAlaThrTyrAspSerAsnCySAAlaCySHisGlnProAs 413
DB 693 AGACATCGAGGCGCGGACAGAGCTTCGCTACACCTGCTGCTGCGCGCAGCGTCCGGA 752
QY 413 pGlyLysGlyValPro 418
DB 753 AGGCAAGGCGACCCCG 768

RESULT 10

US-09-252-991A-9169/c
; Sequence 9169, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9169
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1314)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-9169

Alignment Scores:

Pred. No.: 1.67e-06 Length: 1470
Score: 153.00 Matches: 38
Percent Similarity: 49.53% Conservative: 15
Best Local Similarity: 35.51% Mismatches: 48
Query Match: 5.83% Indels: 6
DB: 4 Gaps: 4

US-10-088-045-2 (1-502) x US-09-252-991A-9169 (1-1470)

QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnCySAAlaCySHisGlnProAspGlyLys 415

DB 453 AAGACCGGCGCGCGCTTACGTCGACAACTCGGGCGCTGCGACCGCAGCGCAAG 394
QY 416 GlyValProAsnAlaPheProProLeuAla---AsnSerAspTyrLeuAsnAlaAspHis 434
DB 393 GGCTATGCCGGGCTTCTCCCGGCGCTGCGCGCAGCCGCTGATGACCGGCGAGCCCG 334
QY 435 AlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsn 454
DB 333 ACTGCTGTCGACATCGTGTCAAGGC-----GGCACCCTGCGCGGCCACCCACGAG 280
QY 455 GlnTyrGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471
DB 279 GCACCGTCGAGCTTCACCATGCGCGCTGCGCTGCGGATGACACGACGAGATCGGC 220
QY 472 AsnValIleThrTyrThrLeuAsnSerPheGlyAsnIleGlyGlnIleLeuSerAlaAsp 491
DB 219 GAGTGTCACTCTCATCCGACCACTGGGCGAACCGGCGCAAGCGTCAAGCTCGAC 160
QY 492 AspValAlaLysAlaLysLys 498
DB 159 GAGTTCAGGCGCGCTGCGCAAG 139

RESULT 11

US-09-252-991A-8797
; Sequence 8797, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8797
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8797

Alignment Scores:

Pred. No.: 2.22e-06 Length: 1764
Score: 153.00 Matches: 38
Percent Similarity: 49.53% Conservative: 15
Best Local Similarity: 35.51% Mismatches: 48
Query Match: 5.83% Indels: 6
DB: 4 Gaps: 4

US-10-088-045-2 (1-502) x US-09-252-991A-8797 (1-1764)

QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnCySAAlaCySHisGlnProAspGlyLys 415
DB 845 AAGACCGGCGCGCGCTTACGTCGACAACTCGGGCGCTGCGACCGCAGCGCAAG 904
QY 416 GlyValProAsnAlaPheProProLeuAla---AsnSerAspTyrLeuAsnAlaAspHis 434
DB 905 GGCTATGCCGGGCTTCTCCCGGCGCTGCGCGCAGCCGCTGATGACCGGCGAGCGCG 964
QY 435 AlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsn 454
DB 965 ACCCTCGTGGTCACATCGTGTCAAGGC-----GGCACCCTGCGCGGCCACCCACGAG 1018
QY 455 GlnTyrGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471
DB 1019 GCACCGTCGAGCTTCACCATGCGCGCTGCGCTGCGGATGACACGACGAGATCGGC 1078
QY 472 AsnValIleThrTyrThrLeuAsnSerPheGlyAsnIleGlyGlnIleLeuSerAlaAsp 491

Db 1079 GAGGTGTCATCTTCATCCGACCAAGCTGGGCAACGAGCGCCAGCGTACGCTGAC 1138
Qy 492 AspValAlaLysLys 498
Db 1139 GAGCTCAGCGCGCTGCCAAG 1159

RESULT 12
US-09-252-991A-8910
Sequence 8910, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 8910
LENGTH: 2256
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: unsure
LOCATION: (1015)
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-8910

Alignment Scores:
Pred. No.: 3,266-06 Length: 2256
Score: 153.00 Matches: 38
Percent Similarity: 49.53% Conservative: 15
Best Local Similarity: 35.51% Mismatches: 48
Query Match: 5.83% Indels: 6
DB: Gaps: 4

US-10-088-045-2 (1-502) x US-09-252-991A-8910 (1-2256)
Qy 396 LysAlaGlyLysAlaThrTyrAspSerAsnCyseAlaIaCysHisGlnProAspGlyLys 415
Db 1876 AAGACGCGCGCGGCTCTACGTGACACATGCGGCGCGCCACCCGACGCGGCAAG 1935
Qy 416 GlyValProAsnAlaPheProPheLeuAla---AsnSerAspTyrLeuAsnAlaAspHis 434
Db 1936 GGCTATGCGCGGCTCTCCCGCCCTGCGCGCAACCGGTAGTACCGGCGGACGCCG 1995
Qy 435 AlaArgAlaIaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAlaAsnGlyAsn 454
Db 1996 ACCCTCGGTGTCACATGCTGCTCAAGGC-----GACACCTCGCGCGGCGGCGGACCCAG 2049
Qy 455 GlnTyrGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471
Db 2050 GCACCGTCGAGCTTCACCATGCGCGCTGCGGCTGGCGGATGAACGACGAGATCGCC 2109
Qy 472 AsnValIleThrTyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAsp 491
Db 2110 GAGGTGTCACATCTTCATCCGACCAAGCTGGGCAACGAGCGCGCCAGCGTACGCTGAC 2169
Qy 492 AspValAlaLysLysLys 498
Db 2170 GAGCTCAGCGCGCTGCCAAG 2190

RESULT 13
US-07-985-458-2
Sequence 2, Application US/07985458
Patent No. 5344777
GENERAL INFORMATION:
APPLICANT: Tamaki, Toshimi;
APPLICANT: Takemura, Hiroshi;

APPLICANT: Tayama, Kenji;
APPLICANT: Fukaya, Masahiro;
APPLICANT: Okumura, Hajime and
APPLICANT: Kawamura, Yoshiya
TITLE OF INVENTION: Structural Gene of Membrane-Bound
TITLE OF INVENTION: Alcohol Dehydrogenase Complex, Plasmid
TITLE OF INVENTION: Containing The Same And Transformed Acetic Acid
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Frischauf, Holtz, Goodman & Woodward, P.C.
STREET: 600 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10016-2088
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 0.72 mb
COMPUTER: IBM PC compatible (NEC PC-9801 ES)
OPERATING SYSTEM: MS DOS
SOFTWARE: ASCII Form
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/985,458
FILING DATE: 19921203
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/658,221
FILING DATE: 20-FEB-1991
APPLICATION NUMBER: 73440/1990
FILING DATE: 26-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Goodman, Herbert
REGISTRATION NUMBER: 17061
REFERENCE/DOCKET NUMBER: 910134/HG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)972-1400
TELEFAX: (212)370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1404 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Acetobacter alioacetigenes
STRAIN: MH-24
PUBLICATION INFORMATION:
AUTHORS: Tamaki, Toshimi;
AUTHORS: Fukaya, Masahiro;
AUTHORS: Takemura, Hiroshi;
AUTHORS: Tayama, Kenji;
AUTHORS: Okumura, Hajime;
AUTHORS: Kawamura, Yoshiya;
AUTHORS: Nishiyama, Makoto;
AUTHORS: Horinouchi, Sueharu and
AUTHORS: Beppu, Teruhiko
TITLE: Cloning and Sequencing of the Gene Cluster
TITLE: Encoding Two Subunits of Membrane-Bound
TITLE: Alcohol Dehydrogenase from Acetobacter
TITLE: polyoxogenes
JOURNAL: Biochimica et Biophysica Acta.
VOLUME: 1088
PAGES: 282-300
DATE: 1991
US-07-985-458-2

Alignment Scores:
Pred. No.: 1,986-06 Length: 1404
Score: 152.00 Matches: 108
Percent Similarity: 38.02% Conservative: 57
Best Local Similarity: 24.88% Mismatches: 165

Query Match: 5.79% Indels: 104
 DB: 1 Gaps: 23
 US-10-088-045-2 (1-502) x US-07-985-458-2 (1-1404)

QY 113 ProGlyIleMetIleArgValArg-----GluGlyAspThrIleGluValGln 128
 DB 156 CCGTGAACATCTACGACGCGCTTGAAATCAAGAGCCCATCGTACGATCTACTC 215
 QY 129 PheSerAsnHis-----ProAsp----- 134
 DB 216 CACGAACATACACCGGACCCGACCTTACGGTATCGCTTACCACTTCCCGCAATTGCA 275
 QY 135 ---SerIleMetPro-----HisAsnValAspPheHisAlaAla 146
 DB 276 CGAAGCGGTGCGCATGTATCCGCAAGACGAGTTCACGCTGTATCCGCGCATGCCGTA 335
 QY 147 Thr-----GlyProGlyGlyValGluAlaSerPheThrAla 159
 DB 336 TCCCTCTCTTCCGCGATGACGAAGAGACATGACGCGCTGTATGCGTATCTCATGCA 395
 QY 160 ProGlyHisThrSerPheSerPheValAlaLeuGlnProGlyLeuValTyrHis 179
 DB 396 TGGGGTGAAGCCGCTGCGCGACCGACAGACGCGCATCTCTGCGCTTGTTCAT 455
 QY 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleVal 199
 DB 456 GCGCTGCGCGCTGG-----CATCTGCGCATGAT-----GTTCTGCC 494
 QY 200 -GluProGlyGlyLeuProGlyValAspIleGlyTyrTyrValMetGlnIleAsp 219
 DB 495 TTGCGCGAAGACCTTACGCGCGCGACGACGCGATCTGAAATCGCATGCGCGCATTA 554
 QY 219 eTyrThrIleGlyLeuTyrGlyGlnGlnIleLeu-----GlnProPheAsp 235
 DB 555 TCTGTTACCGGCGCC---GGGCAATTGCGGTGCGTGCATATCCCGCGTGGCTTCCGAT 611
 QY 235 ---GluTyrAlaIleArgGluAspIleGluTyrValIlePheAsnIleSerVal-Gly 254
 DB 612 GCAGAGAAAGCGCTGACGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 671
 QY 254 ILeuThrGlyGluAsnAlaLeuAlaValAlaGlyIleThrValArgLeuPheValG 274
 DB 672 CGACACTGGGTGCGCGCGCGACCTGCGCA-----ACGATCTCTGCTGCTGCT 719
 QY 274 IAsnIleGlyProAsnLeuThrSer---SerPheHisValIleGlyIlePheAsp 293
 DB 720 GGGCGCGCTGCTGAGATGACATTAACACTTCTGAACTCGCGCGTATGACCACTC 779
 QY 293 sValHisPheGluGlyGlyLeuGluAsnHisAsnIleGlnThrThrLeuIleProAl 313
 DB 780 CCGCGCTTCT---GGTGGCATGGGCGAT----- 804
 QY 313 aGlyGlyAlaIleThrGluPheValAspValProGlyAspTyrValIleValAs 333
 DB 805 ---GTGGTGCATGAGACACCGAGTACTTACCGATGACGACCTG----- 846
 QY 333 pHisAlaIlePheArgAlaPheAsnIleGlyAlaLeuGlyIleLeuValGluGly 353
 DB 847 -CACGCGCATCGCAGAG---TACCTGAAGACCTGCGCGCGCTGCGCGCGCGCA 902
 QY 353 uGluAsnHisGluIleTyrSerHisGlyIleThrAspAlaValTyrIleProGlyGly 373
 DB 903 C-----TACACTTACATCCGCTCACCGCGCAACATGCTGCTGCGGT-- 945
 QY 373 aProGlnAlaIleAspThrGlnIleAlaProGlySerProAlaProAlaAsnLeuGln 393
 DB 946 ---AATACCGCCAGCGTTCCG----- 963
 QY 393 uGlnIleValGlyValAspAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnPro 413
 DB 964 -----GGTCTGATACGTATGTGAAGAAAGGCGCATCTGTACCGTAAACA 1010

QY 413 pGlyIleGlyValProAsnAlaPheProIleuAlaAsnSerAspTyrLeuAsnAla 433
 DB 1011 CCGTGTGGCGTGGCGCGCATGTCTCCGCGCTGCGCAACCCGCTGTCTGACCGCA 1070
 QY 433 pHisAlaAlaGlyAla---AlaSerIleValAlaAsnGlyLeuSerGlyValIleThrVal 452
 DB 1071 GAACCCGACCTGCGCTGTGTAACGTATGCGCATGTGT-----GGGTGCTGCGCGGAG 1124
 QY 452 nGlyAsnGlnTyrGluSerVal---MetProAlaIle-----AlaLeuSerAspGln 469
 DB 1125 CAACCTGGACCGCTCCGCGACGTGCAATGCGGTTACAGCACTGCTGCTGCGCGCA 1184
 QY 469 nIleAlaAsnValIleThrTyrThrLeuAsnSerPheGlyAsnIleGly---GlyLeu 488
 DB 1185 GATTGCTGATGTGTCACTTATCCGACACGACTGGGCGCAACAGCGCGCACCGCT 1244
 QY 488 uSerAlaAspPheValAlaValAlaValAspIleThr 499
 DB 1245 TACGCTGCGGATGTACCAAGCTGCGGACAGC 1278

RESULT 14
 US-09-252-991A-765
 ; Sequence 765, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252, 991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074, 788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094, 190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 765
 ; LENGTH: 1335
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-765

Alignment Scores:
 Pred. No.: 2,066-06 Length: 1335
 Score: 151.50 Matches: 87
 Percent Similarity: 31.83% Conservative: 40
 Best Local Similarity: 21.80% Mismatches: 110
 Query Match: 5.77% Indels: 162
 DB: 4 Gaps: 18

US-10-088-045-2 (1-502) x US-09-252-991A-765 (1-1335)

QY 132 HisProAspSerIleMetProHisAsnValAspPheHisAlaIleThrGlyProGly 151
 DB 528 CACCTGTAT-----CCACATCTACGACACTTCCGAGCCGCGACCTGACGAGT 578
 QY 152 ---GlyAlaGluAlaSerPheThrAlaProGlyHisThrSerPhe 168
 DB 579 CACCGGCTACCACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 617
 QY 169 LyAlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHis 188
 DB 618 -----CSAGAGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 659
 QY 189 IleAlaAsnGlyMetTyrGlyLeuIleLeuValGluProGlyGlyIlePro-LysVal 208
 DB 660 ---CCAGATCCACAA-----CCGACATCCACAA-----CCGACATCCACAA 680
 QY 208 IAspIleGlyTyrTyrValMetGlnIleAspPheTyrThrIleGlyTyrGlyGly 228
 DB 681 GAAGGACGAGCATTACTCTGAGAGTGAAG----- 711
 QY 228 nGlyLeuGlnProPheAspMetGluValAlaIleArgGluAspIleGlyTyrValIlePhe 248

```

Db      711 ----- 711
Qy      248 eAeNglySeRvAlGlyAlaLeuThrglyGluAsnAlaLeuValaLysValGlyGluTh 268
Db      712 ----- 740
Qy      268 rValaLgLeuPheValGlyAeNglyGlyProAsnLeuThrSeRSeRpheniSValIlegl 288
Db      741 GGTGGCTTCCTGATC-----ACCTCCAGC----- 765
Qy      288 yGluilePheAerlySValHisPheGlyGlyLysGlyGlyLysAsnHisValIleglTh 308
Db      766 -----GACGTGATCCAT-----TC 779
Qy      308 rThrLeuileProAlaGlyAla-----AlaileThrgluPhe----- 321
Db      780 CTGTGGTGGCTGGCTTCCTGGGCTCAAGCCGAGCCATCCCGGCTTCGTCACAGAGCC 839
Qy      322 -----LysValaerValProGlyAerTyT----- 329
Db      840 CTGACCAAGGTGCAGAGCCCGCATCTATCGCGCAAGTGGCCGAGGTGGCGCA 899
Qy      330 -----ValleuValaerHisAlailePheArgAlaPheAsnLysG 343
Db      900 GAGCAGCGCTTCATGCCGATCGTGTGAC----- 930
Qy      343 yAlaLeuGlyileuLysValGlyGlyGluAsnHisGlyileTySerHisLysG 363
Db      931 -----GTCAAGCCCAAGCCGAGTTCGACAGTGGCTGGCCCAAGCCCAAGGA 977
Qy      363 nThrAspAlaValTyLLeuProGlyAlaProGlnAlaileAerThrglnGlyAlaPr 383
Db      978 A---GAGCGCGCAAGGTCAAGAA-----CTGACCAAGCAAGAGTGGAC 1019
Qy      383 oLyThrProAlaProAlaAsnLeuGlnGlnIleLysValGlyLysAlaThrTyas 403
Db      1020 CAAG-----GAAGAGTGGTGGCGCGCGCAAGGTCTACCA 1058
Qy      403 pSerAsnCyAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAlaPheProPr 423
Db      1059 CACCATTCGGCGCCCTGCCCAAGCCCAAGCCCAAGCCCAAGTGTTCGCCGC 1118
Qy      423 oLeuAlaAsnSerAerTyLLeuAsnAlaAspHisAlaArgAlaLaserileValaLys 443
Db      1119 GCTGAAGCGTTTGAAGATCGTCAAGCGGCCCAAGACACCACTGGAAGGTCTTCA 1178
Qy      443 nGlyLeuSerGly---LysileThrValaAsnGlyAsnGlnTyGlySerValaLysProAl 462
Db      1179 CGGCGTCCCGGCGCCGACATGGCGGCTTCGCGACAGC----- 1218
Qy      462 aileAlaLeuSerAerGlnGlnIleAlaAsnValileThrTyThrLeuAsnSerPheG 482
Db      1219 -----CTCAACGAGTTCACCTGGCGCGGTGATACCTACAGCGCAACGCTGGGG 1271
Qy      482 yAsnLysGlyGlyLysLeuSerAlaAspValaLysAlaLysValaLysThryS 500
Db      1272 CAAGCAGATGGGACATGATCAAGCCCAAGAGCTGTGCTCAAGAGCAAA 1326

```

```

RESULT 15
US-09-252-991A-737
; Sequence 737, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: MARC J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 737
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-737

Alignment Scores:
Pred. No.: 2,54e-06 Length: 1524
Score: 151.50 Matches: 87
Percent Similarity: 31.83% Conservative: 40
Best Local Similarity: 21.80% Mismatches: 110
Query Match: 5.77% Indels: 162
DB: Gaps: 18

US-10-088-045-2 (1-502) x US-09-252-991A-737 (1-1524)
Qy      132 HisProAspSerLysMetProHisAsnValaAspPheHisAlaLathrglyProGly 151
Db      436 CACCTGAT-----CCACATCTACGACACTTCGAGCCGAGCTGACGTGACAGGT 486
Qy      152 -----GlyAlaGlyLaserPheThrAlaProGlyHisThrSerThrPheSerPhe 168
Db      487 CACCGGCTACCACTGGAAAGTGCAGTACAGTACCTGG----- 525
Qy      169 LysAlaLeuGlnProGlyLeuTyValTyHisCysAlaValaLarProValGlyMetHis 188
Db      526 -----CCAGAGCTGAGTACTTCGACCAACCTGGCCCAAGCCCAAGGA----- 567
Qy      189 IleAlaAsnGlyMetTyGlyLeuileuValaGlnProLysGlyGlyLeuPro-LysVa 208
Db      568 -----CCAGATCCACAA-----CCGAGCAGGC 588
Qy      208 LAspLysGlyTyTyValaMetGlnLysAerPheTyThrLysGlyLysTyGlyGlyG 228
Db      589 GAAGCAGAGCATTACTCTGTGAGGTGAC----- 619
Qy      228 nGlyLeuGlnProPheAspMetCylusAlaileArgGluAspAlaGlyTyValaLPh 248
Db      619 ----- 619
Qy      248 eAeNglySeRvAlGlyAlaLeuThrglyGluAsnAlaLeuValaLysValGlyGluTh 268
Db      620 -----GAGCGCGTGGTGGCGCGCGCAAGGTGGCGCAAA 648
Qy      268 rValaLgLeuPheValGlyAeNglyGlyProAsnLeuThrSeRSeRpheniSValIlegl 288
Db      649 GGTGGCTTCCTGATC-----ACCTCCAGC----- 673
Qy      288 yGluilePheAerlySValHisPheGlyGlyLysGlyGlyLysAsnHisValIleglTh 308
Db      674 -----GACGTGATCCAT-----TC 687
Qy      308 rThrLeuileProAlaGlyAla-----AlaileThrgluPhe----- 321
Db      688 CTGTGGTGGCTGGCTTCCTGGGCTCAAGCCGAGCCATCCCGGCTTCGTCACAGAGCC 747
Qy      322 -----LysValaerValProGlyAerTyT----- 329
Db      748 CTGACCAAGGTGCAGAGCCCGCATCTATCGCGCAAGTGGCCGAGGTGGCGCA 807
Qy      330 -----ValleuValaerHisAlailePheArgAlaPheAsnLysG 343
Db      808 GAGCAGCGCTTCATGCCGATCGTGTGAC----- 838
Qy      343 yAlaLeuGlyileuLysValGlyGlyGluAsnHisGlyileTySerHisLysG 363
Db      839 -----GTCAAGCCCAAGCGAGTTCGACAGTGGCTGGCCCAAGCCCAAGGA 885
Qy      363 nThrAspAlaValTyLLeuProGlyAlaProGlnAlaileAerThrglnGlyAlaPr 383
Db      886 A---GAGCGCGCAAGGTCAAGAA-----CTGACCAAGCAAGAGTGGAC 927

```



```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4893
; LENGTH: 1554
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4893

```

Alignment Scores:

```

Pred. No.: 3,77e-06 Length: 1554
Score: 150.00 Matches: 44
Percent Similarity: 48.61% Conservative: 26
Best Local Similarity: 30.56% Mismatches: 61
Query Match: 5.71% Indels: 13
DB: Gaps: 5

```

US-10-088-045-2 (1-502) x US-09-252-991A-4893 (1-1554)

```

QY 361 HisLysGlnThrAspAlaValTyrLeuProGlu-GlyAlaProGlnAlaIleAspThrG1 380
DB 835 CATGCCAG-----CTACTGAAGTCCCTGCGCGCGCAAGACGACCT 879
QY 380 ngLualPro-----LysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaG1 398
DB 880 GCCCATGCCCGACGACGCGCGCTGCGACGCGCTGACCTGACGCTGCGCGG 939
QY 398 yLysAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValPr 418
DB 940 CGGCTCGGCTACCGCGAGTTCCTGCTCCGACTGCCACCGCAAGACGCGCGCTCCC 999
QY 418 oAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla1 438
DB 1000 GGGCATGTTCCCGCGCTGCGCGCGCAACCCACGCTGCTTCG-----GCCAACCGGAG 1053
QY 438 aSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnIlyrG1use 458
DB 1054 CAGCTACTGCATATACCTGACCGGCTGGAACCGCGGAAACCGCAACCGCATCTCGC 1113
QY 458 rVal-----MetProAlaIleAla--LeuSerAspGlnGlnIleAlaAsnValIleTh 475
DB 1114 GGTCTACACCATGCGCGCTTCCGCGCGAGAGACCGGAAATGCCAGATCTCTCAG 1173
QY 475 rTyrThrLeuAsnSerPheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaL 495
DB 1174 CTTCGTCGCCAGAGTGGGCAACGAGGTTCTGATCGATCGCGCGCGAGTGAAAGA 1233
QY 495 sAlaLysLys 498
DB 1234 ACTGCCGCGAG 1243

```

RESULT 18

```

US-09-252-991A-4910
; Sequence 4910, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4910
; LENGTH: 2139

```

```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4910

```

Alignment Scores:

```

Pred. No.: 6.22e-06 Length: 2139
Score: 150.00 Matches: 44
Percent Similarity: 48.61% Conservative: 26
Best Local Similarity: 30.56% Mismatches: 61
Query Match: 5.71% Indels: 13
DB: Gaps: 5

```

US-10-088-045-2 (1-502) x US-09-252-991A-4910 (1-2139)

```

QY 361 HisLysGlnThrAspAlaValTyrLeuProGlu-GlyAlaProGlnAlaIleAspThrG1 380
DB 942 CATGCCAG-----CTACTGAAGTCCCTGCGCGCGCAAGACGACCT 986
QY 380 ngLualPro-----LysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaG1 398
DB 987 GCCCATGCCCGACGACGCGCGCTGCGACGCGCTGACCTGACGCTGCGGG 1046
QY 398 yLysAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGlyValPr 418
DB 1047 CGGCTCGGCTACCGCGAGTTCCTGCTCCGACTGCCACCGCAAGACGCGCGCTCCC 1106
QY 418 oAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla1 438
DB 1107 GGGCATGTTCCCGCGCTGCGCGCGCAACCCACGCTGCTTCG-----GCCAACCGGAG 1160
QY 438 aSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnIlyrG1use 458
DB 1161 CAGCTACTGCATATACCTGACCGGCTGGAACCGCGGAAACCGCAACCGCATCTCGC 1220
QY 458 rVal-----MetProAlaIleAla--LeuSerAspGlnGlnIleAlaAsnValIleTh 475
DB 1221 GGTCTACACCATGCGCGCTTCCGCGCGAGAGACCGGAAATGCCAGATCTCTCAG 1280
QY 475 rTyrThrLeuAsnSerPheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaL 495
DB 1281 CTTCGTCGCCAGAGTGGGCAACGAGGTTCTGATCGATCGCGCGAGTGAAAGA 1340
QY 495 sAlaLysLys 498
DB 1341 ACTGCCGCGAG 1350

```

RESULT 19

```

US-09-252-991A-8339
; Sequence 8339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8339
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8339

```

Alignment Scores:

```

Pred. No.: 1.01e-05 Length: 1977
Score: 147.50 Matches: 85
Percent Similarity: 34.99% Conservative: 56
Best Local Similarity: 21.09% Mismatches: 147

```


Query Match:	5.62%	Indels:	115
DB:	4	Gaps:	20

Oy	107	ThrPheGlyGlyValProGlyValMetIleArgValArgGlyGlyAspThrIleGlu	126
Db	328	ACCATCAACGGTTCCGCTCCGACCGCAACCTCCGCTGGCGGAAAGCGCAACCTGAC	387
Oy	127	ValGlnPheSerAsnHis-----ProAspSerLys-----Met	137
Db	388	CTGGCGGGTGGCGAACCGGCTCGCGAGACACTGGATCCATCGACAGCGCATTCCTGG	447
Oy	138	ProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluAlaSerPhe	157
Db	448	CCGGCGCAACATGGAT-----GACGTGGCGGGCTTCAGCTTCGAGGGC-----	489
Oy	158	ThrAlaProGlyHisIleThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuThrVal	177
Db	490	ATCGCTCTGGCGGGCTCTCAAGATACCGCTTCAGAGTCCGGCAACAGCACTTACGG	549
Oy	178	TyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIle	197
Db	550	TACCAAC-----AGCCACTCGGGGCTCAAGAGCAAGCGCGGGGTCTACGCGCCCTG	600
Oy	198	LeuValGluProLysGlyGlyLeuPro---LysValAspLysGlyTyrTyrValMetGln	216
Db	601	GTGATTCGACGGCCCGGAGCCGGAAACGTTCAAGTTCAGTCGACCGCGCATCTAGTGTGCTGC	660
Oy	217	GlyAspPheTyrThrIleGlyLysTyrGlyGlyGlnGlyLeuGlnProPheAspMetGlu	236
Db	661	AGCAGCTGCTGGACGACG-----AAGCCCGACGGCATCTCTC	696
Oy	237	LysAlaIleArgGlyAspAlaGlyTyrValValPhe-----	248
Db	697	GCCAAAGCTGAAGAAAGCACTCGCATTAACAATCTCCAAAGCGCACCGTCCGCGCATTC	756
Oy	249	-----AsnGlySerValGlyAlaLeuThrGlyGluAsn-----	259
Db	757	ATCGACGACGTGACGGCCAAAGCGCTGGGCGCGGACCTGGCGGACCGCAAGATGTGGCC	816
Oy	260	AlaLeuLys-----AlaLysValGlyGlyThrValArgLeuPheVal	273
Db	817	GAGATGAAGATGAGTCCACCGCATCTCGCCAGCGTCACGCGGTACACTATACCTGACTG	876
Oy	274	GlyAsnGlyGlyPro-----AsnLeuThrSerSerPheHisValIleGlyGluIle	290
Db	877	CTCAACGGACACCGCGCGGACGCGCAATGACCGCGCTGTTCCGC-----	921
Oy	291	PheAspLysValHisPheGlyGlyLysGlyGlyValAsnHisAsnIleGlnThrThrLeu	310
Db	922	-----CCGGCGGAAAGCTCCGCTCGCTTC-----	948
Oy	311	IleProAlaGlyGlyAlaAlaIleThrGluPheLysValAspValProGly-----	327
Db	949	-----GTCAACGCTCGCGCATGAGCATTTTCAGAGTCCGATTCGCCGCTCAAGATG	1002
Oy	328	-----AspTyrValLeuValAspHis-----	334
Db	1003	ACCGTGTGCGCGCGACGACGACGACGCTGACCGCGTCAAGTGCATGAGATGTGGCATC	1062
Oy	335	AlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGlnGlu	354
Db	1063	GCGGTGGCGGACCTTACGACGTG-----ATCGTGAACCGGGCGCGGAGGCG	1110
Oy	355	AsnHisGluIleTyrSerHisLysGlnThrAspAlaValTyr-----	368
Db	1111	GCGTATACCTCTTGGCGCGAGTGCATGACCGCGAAGGGTTACGCGCGGACACTGGGG	1176
Oy	369	LeuProGlyGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaPro	388
Db	1171	CTGGCGGAAGGCTTAGCGCGCGCGT-----CCAGCGCCGATCCG	1212

Oy	Alahenleunginglunginilelyslagilyvalathrtryapsseraemcyalaala	408
Dd	CGCCCCGTGATGCCATGGACGACATGGCATGGCCGCGCATGC-----	1257
Oy	Cyshtisglnproabpgilylsgilyvalproabnla-----PhePro	422
Dd	---CACGGCGCCATGGGGCATGGCGCGCAGCGCCCGACGAGATGCACCATTCG	1314
Oy	423 Proleuhlahenserrapyryrleuasnalalasphtsalatrgalaalasertlevala	442
Dd	1315 AAGATGTCGGGCATGACATGAAGGCGCATGCATTCGAAAGTGGCGCGCATGGACATCG	1374
Oy	Asngilylen	445
Dd	1375 AACGGCATG	1383

```

RESULT 20
US-09-252-991A-14102
: Sequence 14102, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 31142
: SEQ ID NO 14102
: LENGTH: 1782
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14102

```

Alignment Scores:	
Pred. No.:	1,246-05
Score:	146.00
Percent Similarity:	33.7%
Best Local Similarity:	22.1%
Query Match:	5.56%
DB:	.4
	Gaps: 25
	Indels: 128
	Mismatches: 219
	Conservative: 61
	Matches: 116
	Length: 1782

```

US-10-088-045-2 (1-502) x US-09-252-991A-14102 (1-1782)

Oy      68 HisAlaProGluValProProProValAspArgAspHisSerAlaLysValValLys 87
      |||
      274 CATCGCGAGCTGAAACCCGATCCAGACAAAGTTCTGTGACAGCGATTCCCTCCAGTCGAC 333
      |||
Oy      88 MetGluThrValGluLysValMetArgLeuAlaSerGluValGluLysGlnPheThrPThr 107
      |||
      334 CTA-----TTGCGCGCGCGGCTTCGTCAAGCGCGCCACCGT 369
      |||
Oy      108 PheGluGluGlnValProGly-----GlnMetLeaArgValArgGluGluAspThr 124
      |||
      370 GCTGTGTAATAAAGCCAGCCAGCCGCGCATCAAGCGACGACGACCTGGACAGCGCATCGA 429
      |||
Oy      125 TleGluValGlnPheSerAspHisProAspSerLysMetProHisValAspPheHis 144
      |||
      430 GAGCAGCGCTCGGCATCATCTGCGCTTGACACAGCGATCGTGCCTACTACAGCGCCAC 489
      |||
Oy      145 -----AlaAlaThrGlyProGly---GluGluValGluAlaSerPheThrAla 159
      |||
      490 CCGCGACGTGTCCAGACGAACCTGCGCTGTGTCAAGAGGCTTGAGATGAAGCGACGACCTT 549
      |||
Oy      160 ProGluHisThrSerThrPheSerPheLysAlaLeuGlnPro----- 173
      |||
      550 TCCGGC-----CTCGCGCTGGCGCTGGCGCGGCTGTGCGCGCTTCCTGGCGCTCGCGGCC 603
      |||
Oy      174 -----GluLeuTyrValTyrHis----- 179

```

```

Db      604 GGCAGCCGAGCGGCTGTGTCAGCGGCGGAGATACCTCCGCCGCCGAGCTGCATG 663
Qy      180 ---CysAlaValAlaProValGlyMetHisLeuAlaAsnGlyMet----- 193
Db      664 GCCTGCCATACCGGAGGCGGCGCGCTTCCGCCGCGGCGCTGCGGATTCAGTCCGCC 723
Qy      194 -----TyrGlyLeuLeuValGluProGlyGlu-----GlyLeuPro 206
Db      724 TTGGGACCATCTACGGGACCAACATCAACCCGAGACAGAGAACGGGATCGGCGCTAC 783
Qy      207 LysValAspLysGlyTyrTyrValMetGlnLysAspPheTyrThrLysGlyLysTyrGly 226
Db      784 AGCGCCACGAGTTCCTCGCGCTCTT-----ACCGAGGCGAGCGGAG 828
Qy      227 GluGlnGly-----LeuGlnProPheAspMetGluValAlaIle---ArgGlu 241
Db      829 GAGCGCCCTACCTCTATCCGCGCATGCCCTACCTCTATCACTGATCGAGCGGAG 888
Qy      242 AspAlaGlyTyrValVal-----PheAsnGlySerValGlyAlaLeuThrGlyLysAla 260
Db      889 GATGCCGACCGCATCTACGCTTACCTGATGCGCCAGAGCCGATCGCCGCCGCCG 948
Qy      248 -----PheAsnGlySerValGlyAlaLeuThrGlyLysAla 260
Db      949 CAGACGAGCTGAGCTTCCGCTTCACTGCGCATGGGC---CTGGCCGCGTGGAACTG 1005
Qy      261 LeuLysAlaLysValGlyLysValThrValArgLeuPheValGlyLysGlyProAsnLeu 280
Db      1006 CTC-----TAGCGAAGAGGCTGGCTTGCAGCGCGAGAGAGGAGCGGAGAGCC 1056
Qy      281 ThreSerPheHisValIleGlyLysLeuPheAspLysValHisPheGlyGlyLys 300
Db      1057 TGGAAAGCGCGGAGCATGATGCTGAGAGTCTC-----GGCCACTGC 1098
Qy      301 GlyLysAsnHisValIleGlnThrLeuLeuProAlaGlyAlaAlaIleThrGlu 320
Db      1099 GCGAGAGCGCATACCCGCGGAC---CTGGCGCGCGCTGGAGCAGAGAC 1146
Qy      321 PheLysValAsp-----ValProGlyAsp 328
Db      1147 AAGCGCTGAGCGGCGGCTGCTCAAGCGCTACCTGGCGCGGAGCTGCTCCGCGGAC 1206
Qy      329 -----TyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGly 343
Db      1207 CTGGCGCGGCGGCTTGAACCCAGCGGACCTGCGAGCTTCTCAAGACGCGATAGC 1266
Qy      344 AlaLeuGlyLysLeu-----LysValGlyLysGluAsnHisGlu 357
Db      1267 GCCCAGGCGAGCATGTTCAACGAGATGTTCCGCTGGTGCACACAGCAGCAGCCTC 1326
Qy      358 IleTyrSerHisLysGlnThrAspAlaValTyr---LeuProGlyLysAlaProGlnAla 376
Db      1327 GACGACAGCAGCTGCGAGCCATGCCATCTGCTGCGGCGAGCGCGCGCGCG 1386
Qy      377 IleAspThrGlnGluAlaProLysThrProAlaProAlaAsnGlnGlnIleLys 396
Db      1387 AAGTGTCTCAGCGCGCTCCGAG-----GCCAGCTGAACGACAGCGGAG 1434
Qy      397 AlaGlyLysAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLysGly 416
Db      1435 CGCGGTGCGCAGCATCTCAACGCTGCGCGCGCTGCCACGCGCGGAGGAGGCG 1494
Qy      417 ValProAsn---AlaPheProLeuAlaAsnSerAspTyrLeuAlaAlaAspHisAla 435
Db      1495 AAGCGCATATCGCGTGGCGATGAAACGCGCAACTACCTCGCGCTGCAAGACCGCGT 1554
Qy      436 ArgAlaLysSerIleValAlaAsnGlyLysSerGlyLysIleThrValAsnGlyLysGln 455
Db      1555 AATCTGTTGGAGATCGAAGAGCGATCTCAACAG---CAGTTCCACGCGCTTGAA 1611
Qy      456 TyrGluSerValMetProAlaIleAla---LeuSerAspGlnGlnIleAlaAsnVal 473

```

```

Db      1612 CGCATGCAACCGATGCCGGTTTCGCCGGGAACTGACAGCAGCACTACGAGCTG 1671
Qy      474 IleThrTyrThrLeuAsnSerPheGlyValLysGlyGlnLeuSerAlaAspVal 493
Db      1672 CTCACTACTCTGGCGGACAGACTGGGCGGATTCGCCGGTACCTCGGTCCCAACAGT 1731
Qy      494 AlaLysAlaLys 497
Db      1732 GCGCAATTGAA 1743

RESULT 21
US-08-706-037-24
Sequence 24, Application US/08706037
Patent No. 5770419
GENERAL INFORMATION:
APPLICANT: Xu, Feng
APPLICANT: Betka, Randy M.
APPLICANT: Wahleithner, Jill A.
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57704190 No. 5770419disk of No. 5770419th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,037
FILING DATE: 30-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,800
FILING DATE: 1-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4526.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1588 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1587
US-08-706-037-24

Alignment Scores:
Pred. No.: 3,49e-05 Length: 1588
Score: 141.00 Matches: 103
Percent Similarity: 32.28% Conservative: 60
Best Local Similarity: 20.40% Mismatches: 216
Query Match: 5.37% Indels: 126
DB: 1 Gaps: 20

US-10-088-045-2 (1-502) x US-08-706-037-24 (1-1588)
Qy      59 LeuProValIleAspAlaIleValThrHisAlaProGluValProProProValAspArg 78
Db      22 CTACCTTCTGCTGCGGTCTCAAC----- 48

```

OY	79	AspHisProAlaValValValValys-----MetGluThrValGluValMetArg	96
Db	49	-----CCGGCTTGGCTTCCCTCCGCAACTATTAAGTTGCATCAAGAACTCAATGTC	102
OY	97	LeuAlaAspGlyValGluIuTrgInPheTrpThrPheGlyGlyValValProGlyGlnMet	116
Db	103	GCTCCCGAATGGCTTTCAGCGCTTATCGTCCGTCMAAGGTTTATGTTCTTCGACAGTTG	162
OY	117	ILeaGValArgGluGlyAspThrIleGluValGlnPheSerAsnHis-----Pro	133
Db	163	ATCAGCGCCMAAGGATGACACTTGGCAATTAATGTCACGAAATCACTCAGCGACCT	222
OY	134	AspSerLysMetProHisAsnValAspPheHis-----AlaIleThrGlyPro	149
Db	223	AGTATGCGCTCGTGGCCACAAACGATTCATTGGCATTCGATTGTTCCAAAGTCTACAGCCGAC	282
OY	150	GlyGlyGlyAlaGluAlaSerPheThrAlaPro-----GlyHisThrSer	164
Db	283	GAGATAGCCCGCGCATTCGTCAACGAAATCCCTATTGGCCAAATTTGCCCTATATACATAC	342
OY	165	ThrPheSerPheLysAlaLeuGlnProGlyLysTrpValTyHisCysAlaValAlaPro	184
Db	343	GAGATCCCATTCGCGCGGCG---CAAAACAGAAACCATGTGATACAC-----GCCCAT	390
OY	185	ValGlyMetHisIleAlaAsnGlyMetTrpGly---LeuIleLeuValGluPro-----	201
Db	391	CTTGGCAAGTCAATATGTCATGATGATGGAGAGCCCTTGGTGCATATATATCAATCCAAACGAC	450
OY	202	-----LysGluGlyLeuProLysValAspLysGluTrpTyTrpValMetGlnLysAspPhe	219
Db	451	CCACACAAAGTGGCGGTACGACGTGATGATGGAGACACAGTATGTCATGTTGACAGACTGG	510
OY	220	Tyr-----ThrLysGlyLysTrpGlyGluGlnGly-----	229
Db	511	TACCATACTCCGGCACCCGTTCTAAGAAAGCAAAATGTTCTGCATATATACACCGCTCTG	570
OY	230	LeuGlnProPheAspMetGlnLysAlaIleArgGluAspAlaGluTrpValValPheAsn	249
Db	571	CTCTCTCCGTTCCGCACTCGGGCTTTATCAATGGCAAGGGCGGCTATATG-----	621
OY	250	GlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValAlaLysValGlyGluThrVal	269
Db	622	-----GGCGGTCCCGCACTCCCGGCTACGTAATCAAGTAAACGTGGAAACGATAT	675
OY	270	ArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGln	289
Db	676	CGCTTGCAGTATACAGCGCTTCTGCTATCGGGTGGTTCACCTTTTCGATC-----	726
OY	290	IlePheAspLysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThrThr	309
Db	727	-----GAAAGCAATAGTCTG-----AAGGACATATGCTG-----ACT	744
OY	310	LeuIleProAlaGlyGly-----AlaAlaIleThrGluPheLysValAsp	324
Db	745	GTCAATTAGGCGCGAATGGATCTCGACCAACGCGCTTGGCTGTTGACAGCTTCCAGATTAC	804
OY	325	ValProGlyAspTrpValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAla	344
Db	805	GCTGGACAACGCTACTCTGTCTATCGTTGAAAGCAACAAACGCGGCAATGATCGAATT	864
OY	345	LeuGlyIleLeuLysValGluGlyGluGluAsnHisGluIleTySerHisLysGlnThr	364
Db	865	CGTGACCAACATGACCGTTGGACAGAACCCGGAACCATATGCAATGGACCCCAACATATGC	924
OY	365	AspAlaValTrpLeuProGluGlyAlaProGlnAlaIleAspThrGlnGlu-----	381
Db	925	TTTGGCCGATTGCACTACGAGGGAAGCGCCCAACGCCGAACCAACGAGAACAGGCAAGT	984
OY	382	-----AlaPro	383
Db	985	GCTATCGGATAGCACTCGTTGAAGAAACCTCAATGGCTCATCAACCTGGCGGCTCGG	104
OY	384	LysThrProAlaProAlaAsnLeuGlnGlnIleLysValGlyLysAlaThrTrpAsp	403

```

Db      1045 GCGCGCTCCGCTCCCGCAGACGTTTCCCTCAATTTGCAATTTGGCCGACGACGTTGAT 1104
Oy      404 Ser-----AsnCYsaAlaAlaCYshisglnProaspGlyLygIyVal 417
Db      1105 GCGATTCCTTAGGTTACATTTTAATACATCAAGTACGAGCGCTCCT----- 1149
Oy      418 ProAsnAlaPheProPoleuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla 437
Db      1150 -----TCGTGCCCCAGCTC----- 1164
Oy      438 AlaSerlleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlu 457
Db      1165 TTGAAGATTTTGGCAACAATGCAGACGAATGACGCCCATTTACACCCAAATGACACACT 1224
Oy      458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThr 477
Db      1225 ATGCTATTGCACACAACTAAAGTTATGAGCTCAATATCAACCGAGGCGACACACCT 1284
Oy      478 LeuAsnSerPheGlyAsnLysGlyGlyGlnLeuSerAlaAspAspValAlaLysAlaLys 497
Db      1285 ATCCATCTCCACGCGCCAT-----GTGTTTGATATGTCATAATCACTC 1326
Oy      498 LysThrLysProAsn 502
Db      1327 GGTGTAACCCGAGAC 1341

RESULT 22
: Sequence 24, Application US/09005397
: Patent No. 5972670
: GENERAL INFORMATION:
:   APPLICANT: Xu, Feng
:   APPLICANT: Berka, Randy M.
:   APPLICANT: Wahlelchner, Jill A.
:   TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
:   TITLE OF INVENTION: ENHANCED ACTIVITY
:   NUMBER OF SEQUENCES: 29
:   CORRESPONDENCE ADDRESS:
:   ADDRESSSEE: NO. 59726700 No. 5972670disk of No. 5972670th America, Inc.
:   STREET: 405 Lexington Avenue, 64th Floor
:   CITY: New York
:   STATE: New York
:   COUNTRY: United States of America
:   ZIP: 10174-6401
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/09/005,397
:   FILING DATE:
:   CLASSIFICATION:
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 08/706,037
:   FILING DATE: 30-AUG-1996
:   APPLICATION NUMBER: US 60/002,800
:   FILING DATE: 1-SEP-1996
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Harrington, James J.
:   REGISTRATION NUMBER: 38,711
:   REFERENCE/DOCKET NUMBER: 4526,200-US
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 212-867-0123
:   TELEFAX: 212-878-9655
:   INFORMATION FOR SEQ ID NO: 24:
:   SEQUENCE CHARACTERISTICS:
:   LENGTH: 1588 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: DNA (genomic)

```

FEATURE:
NAME/KEY: CDS
LOCATION: 1..1587
US-09-005-397-24

Alignment Scores:

Prod. No.: 3,49e-05 Length: 1588
Score: 141.00 Matches: 103
Percent Similarity: 32.28% Conservative: 60
Best Local Similarity: 20.40% Mismatches: 216
Query Match: 5.37% Indels: 126
DB: 2 Gaps: 20

US-10-088-045-2 (1-502) x US-09-005-397-24 (1-1588)

```

QY 59 LeuProVal1LeuAspAla1LeuValThrHisAlaProGluValProProValAspArg 78
DB 22 CTACCTTGGCTCGCTGGCTCAAC----- 48
QY 79 AspHisProAlaValValValValValValValValValValValValValValValVal 96
DB 49 -----CCCGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 102
QY 97 LeuAlaAspGlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyValGly 116
DB 103 GCTCCCGATGGCTTCAAGCCTCTATGCTCCGTCACGCTTAAAGCTTAAAGCTTAAAGCTT 162
QY 117 IleArgValArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 133
DB 163 ATACCGGCAACAGGGTGACCTTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 222
QY 134 AspSerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyPro 149
DB 223 AGTATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 282
QY 150 GlyGlyGlyAlaGluAlaSerPheThrAlaPro-----GlyHisThrSer 164
DB 283 GAGATGGCCCGCATTCGTACACGCAATGCCCTATTCGCAAAATTTGCTCTATACATAC 342
QY 165 ThrPheSerPheLysAlaLeuGlnProGlyLeuValThrHisCysAlaValAlaPro 184
DB 343 GAGATCCCATTCGCGGC-----CAACAGAACCATGTGGTATCC-----GCCCAT 390
QY 185 ValGlyMetHisValAlaAsnGlyMetGly-----LeuIleLeuValGluPro----- 201
DB 391 CTGGCGATCAATATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
QY 202 -----LysGluGlyLeuProLysValAspLysGlyValMetGlnGlyAspPhe 219
DB 451 CCACACAAATCGCGCTACGAGTGAATGCGACAGTACATGATGATGATGATGATGATGATG 510
QY 220 Tyr-----ThyLysGlyLysValGlyGluGluGluGluGluGluGluGluGluGlu 229
DB 511 TACCATATCCCGGACCCGCTTAAAGAAATGTTCTGCACTAAATTAACCCGCTCTG 570
QY 230 LeuGlnProPheAspMetGluValAlaIleArgGluAspAlaGluValValPheAsn 249
DB 571 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 621
QY 250 GlySerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAlaValGlyGluThrVal 269
DB 622 -----GGCGGTCGCGAGTTCCCGGCTCAGTAACTGAACAACTGGGAAACGATAT 675
QY 270 ArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerPheHisValIleGlyGlu 289
DB 676 CGCTTGGCTTAAATCAACGCTTCTGCTATCGGCTGCTTCTGCTTCTGCTTCTGCTTCTG 746
QY 290 IlePheAspLysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThr 309
DB 727 -----GAGGACATTAAGTCTG-----ACT 744
QY 310 LeuIleProAlaGlyGly-----AlaAlaIleThrGluPheLysValAsp 324
DB 324 ----- 324

```

```

DB 745 GTCATTAGACCGCATGAGTCTCTGCACAGCCCTGCTGTTCAGACTTCAGATTAC 804
QY 325 ValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAla 344
DB 805 GCTGGACAAAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 864
QY 345 LeuGlyIleLeuLysValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 364
DB 865 CGTGACCAATGACCGTTGACGAGGCGGACCAATGACCAATGACCAATGACCAATGAC 924
QY 365 AspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGlu----- 381
DB 925 TTTCGCTATTCATGACGAGGAGGCGCCCAAGCGCCCAAGCGCCCAAGCGCCCAAG 984
QY 382 -----AlaPro 383
DB 985 GCTATGCTACTGCTCTGTTAAAGAAACCTCCATGCGCTCAACCTGCGCTCCG 1044
QY 384 LysThrProAlaProAlaAsnLeuGluGluGluGluGluGluGluGluGluGluGluGlu 403
DB 1045 GGGGCTCGCGCTCCCGCAGCGCTTCCCTCAATCTTGCATTTGGGCGCAGCATGTTGAT 1104
QY 404 Ser-----AsnCysAlaAlaCysHisGlnProAspGlyLysGlyVal 417
DB 1105 GGGATCTTAGGTTCACTTAAATTAATCAATCAATGAGAGCTCTCT----- 1149
QY 418 ProAsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla 437
DB 1150 -----TCGTTGCCCGCGCTC----- 1164
QY 438 AlaSerIleValAlaAsnGlyLysSerGlyLysIleThrValAsnGlyAsnGlyVal 457
DB 1165 TTGAAGATTTTGGCAAAATGCGACGATGACGCGGATTTACGCGCAATGACGACACT 1224
QY 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyr 477
DB 1225 ATCGTATTCGACCAATTAAGTATTCAGCTCAATATCAACGAGTGCAGACCACT 1284
QY 478 LeuAsnSerPheGlyAsnLysGlyGlyGluLeuSerAlaAspAspValAlaLysAlaLys 497
DB 1285 ATCATCTCCACGGCCAT-----GTTGTTGATATTCGCAATCACTC 1326
QY 498 LysThrLysProAsn 502
DB 1327 GGTGTACCCGAC 1341

```

RESULT 23
US-08-172-331B-13
Sequence 13, Application US/08172331B
Patent No. 5480801
GENERAL INFORMATION:
APPLICANT: Wahlthner, Jill A.
APPLICANT: Christensen, Bjorn E.
TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5480801 No. 5480801disk of No. 5480801th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/172,331B
FILING DATE: 22-DEC-1993
CLASSIFICATION: 435

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER: US 08/122,230
3      FILING DATE: 17-SEP-1993
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: US 06/122,827
6      FILING DATE: 17-SEP-1993
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: US 08/162,827
9      FILING DATE: 03-DEC-1993
10     ATTORNEY/AGENT INFORMATION:
11     NAME: Lowmye Dr., Karen A.
12     REGISTRATION NUMBER: 31,274
13     REFERENCE/DOCKET NUMBER: 4052.020-US
14     TELECOMMUNICATION INFORMATION:
15     TELEPHONE: 212-867-0123
16     TELEFAX: 212-878-9655
17     INFORMATION FOR SEQ ID NO: 13:
18     SEQUENCE CHARACTERISTICS:
19     LENGTH: 1672 base pairs
20     TYPE: nucleic acid
21     STRANDEDNESS: single
22     TOPOLOGY: linear
23     MOLECULE TYPE: cDNA
24     ORIGINAL SOURCE:
25     ORGANISM: Rhizoctonia lacaseae
26     FEATURE:
27     NAME/KEY: CDS
28     LOCATION: 85..1671
29     US-08-172-331B-13

```

Alignment Scores:

Pred. No.:	3.79e-05	length:	167
Score:	141.00	Matches:	103
Percent Similarity:	32.28%	Conservative:	60
Best Local Similarity:	20.40%	Mismatches:	216
Query Match:	5.37%	Indels:	126
DB:	1	Gaps:	20

US-10-088-045-2 (1-502) X US-08-172-331B-13 (1-1672)

```

Oy      59  IeuProValIleAspAlaIleValThrHisAlaProGluValProProProValAspArg 78
       |||||:::|||||
Db      106  CTACCTTGTGCTCGCGGGCTGCACC-----1323
Oy      79  AspHisProAlaLysValValValLys-----MetGluThrValGluLysValMetArg 96
       |:::|:::|
Db      133  -----CCGCGCTTTGCTGCGCGGCCCAACTAAAGTTGCATGAACAAGCAAGCATGAATGTC 186
Oy      97  LeuAlaAspGlyValGluTyrGlnPheTrpThrPheGluGlyGlnValProGlyGlnMet 116
       |||||:::|:::|
Db      187  GCTCCGAGTAGCTTTCAGCGCTTAAGCTCCTGCAACGGTTAATTGATTCCTGGACAGCTTG 246
Oy      117  IleArgValArgGluGluAspThrIleGluAlaGlnPheSerAsnHis-----Pro 133
       |||||:::|:::|
Db      247  ATCACGGCCAAGAAGGTGACACCTTGGCCATTAAATGTCAGCAATCAACATCAACGAGACCT 306
Oy      134  AspSerLysMetProHisAsnValAspPheHis-----AlaAlaThrGlyPro 149
       :::|:::|:::|
Db      307  AGTAGCGCTGTCSCCAACAAGATTCATTGGCATGATTTGTTCSAAGCTACTAACCGCGGAC 366
Oy      150  GlyGlyGlyValAlaGluAlaSerPheThrAlaPro-----GlyHisThrSer 164
       |||||
Db      367  GAGAGTAGCCCCCGCATTCGTGCACGCAATGCCCTTAATGGCAAAATTTGTCTTAATCATAC 426
Oy      165  ThrPheSerPheLysAlaLeuGlnProGlyLysTyrHisCysAlaValAlaPro 184
       :::|:::|
Db      427  GAGATCCCATTTGGCGCGGC---CAACACAGAACCAATGTGGTATCAC-----GCCCAT 474
Oy      185  ValGlyMetHisIleLeuAlaGlnGlyMetTyrGly-----LeuIleLeuValGlnPro----- 201
       |||||:::|:::|
Db      475  CTTGCGAGTCAATAATGTCCATGTGATTCGACAGGCGCTTTGGTGTATTTAATGATCAAACGAC 534
Oy      202  -----LysGluGlyLeuProLysValAspLysGlyTyrTyrValMetGlnGlyAspPhe 219

```

Db 535 CCACACAAAGTGGCGCTACGACGTGGATGAATGCGGACGACAGTACGACCTTGAAGGACTGG 594
 Oy 220 TTT-----ThryeGlyLysTyrGlyGlnGlnGly----- 229
 Db 595 TACCAATCTCCGGCAGCCGGTTCTAGAAAGGAAATGTTCTGCATTAATTAACACCGCTCTG 654
 Oy 230 LeuGlnProPheAspMetGluLysValIleArgGluAspAlaGluTyrValValPheAsn 249
 Db 655 CTCTCTCTCTGTTCCGGACCTCGGGCTTTATCAATGGCAAGGCGCGTTATGTG----- 705
 Oy 250 GlySerValGlyAlaLeuThrGlnGlyLysAsnAlaLeuLysValLysValGlyGluThrVal 269
 Db 706 -----GCGGCTCCCGCAGTTCCCGGCTCAGTATTAACGTTAAACCTGGGAAACGATAT 759
 Oy 270 ArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlu 289
 Db 760 CGCTTGGCGGTAAACAAACGCTTCTGCATATGGGGCTTTACCTTTTGCACT----- 810
 Oy 290 IlePheAspLysValHisPheGlnGlyGlyLysGlnLysAsnIleGlnThrThr 309
 Db 811 -----GAAAGCACTACTGTG-----ACT 828
 Oy 310 LeuIleProAlaGlyGly-----AlaAlaIleThrGluPheLysValAsp 324
 Db 829 GTCAATTAGGCGGATGGAGATCTTGACACCAACCTTGCGCTGTGACACTTCCCAATTATAC 888
 Oy 325 ValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAla 344
 Db 889 GCTGGACAAACGCTACTCTGTCTATCTGTGAAGCCAAACCAACCGCGCCAACTACTGGATT 948
 Oy 345 LeuGlyIleLeuLysValGlnGlyGlnGluAsnHisGluIleTyrSerHisLysGlnThr 364
 Db 949 CGTCACCAATGACCGTTGCAGGAGCGCGCAACCAATCAACTTGAGACCCACCAATATGC 1008
 Oy 365 AspAlaValTyrLeuProGlnGlyAlaProGlnAlaIleAspThrGlnGlu----- 381
 Db 1009 TTTCGCGATTATGCACTACGAGGAGCGCCCAACCGCCAAACCCAGACGAAACAAGGAGT 1068
 Oy 382 -----AlaPro 383
 Db 1069 GCTATCGGTACTGCATCTGTTGAAGAGAACTCATCGCTCATCAACCTGGCGCTCCG 1128
 Oy 384 LysThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAlaThrTyrAsp 403
 Db 1129 GCGGCGTCCGCTCCCGCAAGCGTTTCCCTCAATCTTGCATTTGGCGCCACAGACGTTGAT 1188
 Oy 404 Ser-----AsnCysAlaAlaCysHisGlnProAspGlyLysGlyVal 417
 Db 1189 GCGATTCTTAGGTTCACTATTTAATTAACATCAAGTACAGGCTCT----- 1233
 Oy 418 ProAsnAlaPheProPheProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla 437
 Db 1234 -----TCGGTGGCCACCGCTC----- 1248
 Oy 438 AlaSerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlu 457
 Db 1249 TTGAAGATTTTGGCAACATGCGAGAAAGAGCGCGCATTTACCGCAAAATGACACAT 1308
 Oy 458 SerValMetProAlaIleAlaLeuSerAspGlnGlnIleAlaAsnValIleThrTyrThr 477
 Db 1309 ATCTATTATGCGACACATTAAGTATGACCTCAATATCCACCGGAGGTGACACCAACCT 1368
 Oy 478 LeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspValAlaLysValLys 497
 Db 1369 ATTCATCTCCACGGCCAT-----GTGTTGTATATGTCGAATATCACTC 1410
 Oy 498 LysThrLysProAsn 502
 Db 1411 GGTCGTACCCCGAAC 1425

```

; Sequence 208, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Blakeberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.10034U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: PaeSeq for Windows Version 3.0
; SEQ ID NO 208
; LENGTH: 980
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-615-192A-208

Alignment Scores:
Pred. No.: 4,91e-05 Length: 980
Score: 136.50 Matches: 71
Percent Similarity: 34.51% Conservative: 46
Best Local Similarity: 20.94% Mismatches: 129
Query Match: 5.20% Indels: 93
DB: 4 Gaps: 14

US-10-088-045-2 (1-502) x US-09-615-192A-208 (1-980)

Qy 78 Argpshisproalalyvalvallylmetcuthrvalglulysvalmetargleu 97
Db 27 CGTAACTACACATTCAATGGTATGAGAAATCAACG-----AGACTG 71
Qy 98 Alaasprlyvalglutryglnphetrpthrphelglyglulvalproglylmetile 117
Db 72 TGTTCG-----ACGAAGCCCATGTGACCTGTAACGGGATGTTCCGGAGCCACTCTC 125
Qy 118 Argvalarggluglyasprthrileglulvalglinpheseranhisproasprserlysmec 137
Db 126 TATGCTAGGGAAGATGACACCGCTGCTGTGAGGCTCTTAAC-----CGTCTC 173
Qy 138 Prohisasnvalasprhehis-----Alaialthr 147
Db 174 AATATACATATGTCACCATTCATGGCATGTATCCGCACTTGACAGCGGGTGGCGAC 233
Qy 148 Glyproglyglulvalglualaserphenralaproglylhisthrserthrpheser 167
Db 234 GGGCCACATACATTACCAATGCCCATC---CAGCCGGGCGAAAGCTATGTATACAT 290
Qy 168 Phelysalaieu---glnproglyleuvaltyrvaltyriscysalavalaprovalgly 186
Db 291 TTGCATCATCAGCGGCAACGGGCAACCTCTGTGGCATGACACATA----- 338
Qy 187 Methislaiealaenglymettyrlyleuileuvalglupolyglygluleu--- 205
Db 339 CTGGGCTCAGGGCAACCCCTGACGAGCCATTGTCTTGTCCCAAGCGTGGTCCA 398
Qy 206 -----Prolyvalasprlysglutrytyrvalmetglnlyasprphetyrthrlys 222
Db 399 TACCTTTCTCTAAACCCACAGAGAGTTGTTGTCTGATTGGCGAATGTGAAATCT 458
Qy 223 Glytyrtyrly-----glulnlyleuaglnprophasprmetglu 236
Db 459 GATCAGAGAGGTGATCAGTCAAGCATCAAGTCCGGAATTAGACACG----- 506
Qy 237 Lyealaileargluaspralaglutryvalvalpheanglyservalglyalaleuthr 256
Db 507 -----AATGTCTCGATGCTCACAGATCAATGGCCATCCAGGGCGCAAGTTCC 554

```

```

Qy 257 -----Glygluaspralaleuvalalysvalglulthrvalargleu 271
Db 555 AATTGCCCTTCCAGAGGTGATTTACGTTGCCGTTGAGAGTGGCAAGATACATGCTG 614
Qy 272 Phevalglyasnlygluproanleuthrserserphenisvalileglulthlepe 291
Db 615 CGAATCATCAAGCGCTCGCTCAATGAGAGAGCTTCTTCAAGATTCCTCCGGGCAC 668
Qy 292 Asprlyvalniehethelulglulysgluluasnhisasnileglnthrthleuile 311
Db 669 -----CAGCTGACCATGTCG 663
Qy 312 Proalaglyglualalalethrngluphelysvalasprvalproglyasprtyrvalleu 331
Db 684 GAGGTGAGCGCCACCTACGTCACAGCTTTCAGACCGCACG----- 725
Qy 332 Valasprhisalalephearalapeasnllysglyalaleuylleuvalglu 351
Db 726 -----ATCGTATTCGACCT 740
Qy 352 Glygluluasnhisglutlyrserhislysglnthralspralaleuvalprogly 371
Db 741 GGCCTAACCAACCAATGCCCTCATCTCCACCGACAGAGCTTGGCAAGTACATGTCGCC 800
Qy 372 Gly-----Alaproglinalaleasprthrnglnlualaprolysthr 385
Db 801 GCCTCCCTTTTATGATGCCCGCATGCCGTGACACATGACCGGACCGGCACA 857

RESULT 25
US-09-252-991A-8410
; Sequence 8410, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8410
; LENGTH: 1203
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8410

Alignment Scores:
Pred. No.: 0.000179 Length: 1203
Score: 132.50 Matches: 45
Percent Similarity: 42.78% Conservative: 32
Best Local Similarity: 25.00% Mismatches: 64
Query Match: 5.05% Indels: 39
DB: 4 Gaps: 8

US-10-088-045-2 (1-502) x US-09-252-991A-8410 (1-1203)

Qy 107 Thrphelglyglulvalproglylmetileargvalargglulysasprthrileglu 126
Db 344 ACATCATCAGCGTTCGCTGCGCCGCGGACCGCTGCGGAGAGCGCAACGTCGAC 403
Qy 127 Valglnpheseranhis-----Proasprserlysmec 137
Db 404 CTGCGGATGCGCAACCGCTCGCGGAGACACCTTCATCCATGTCGACGCGCATATCTTG 463
Qy 138 Prohisasnvalasprhehisalalethrnglyglulvalglualaserphe 157
Db 464 CCGGCCAATCATGAT-----GCGTGTCCGGGCTCTGACCTTCGAGGGC----- 505

```

QY 158 ThrAlaProGlyHisThrSerThrPheSerPheValAlaLeuGlnProGlyLeuTyVal 177
| | | | | : : : | | | | | : : : | | | | |
Db 506 ATGCTCTCCGCGCTCTACGATACCGCTTCAAGGTCCGCGAAGCGCACCCTACTCG 565
QY 178 TyrHisGlyAlaValAlaProValGlyMetHisIleAlaSerGlyMetTyGlyLeuIle 197
| | | | | : : : | | | | | : : : | | | | |
Db 566 TACAC-----ACCCACTCGGGGCTACAGAGAGCGCGGGGCTACGGGCGCTG 616
QY 198 LeuValGluProGlyHisGlyLeuPro---LysValAlaPlyGlyTyTyValMetGln 216
| | | | | : : : | | | | | : : : | | | | |
Db 617 GTATGAGCGCGCGCGAGCGCGAAGCGCTTACAGCTACAGACCGGCTACGCTGCTCTC 676
QY 217 GlyAspPheTyThrTyrGlyTyGlyGlyGlnGlyLeuGlnProPheAspMetGln 236
| | | | | : : : | | | | | : : : | | | | |
Db 677 AGGACTGCTCGGACGAG-----AAGCGCAGCGCATCTCTC 712
QY 237 LysAlaIleArgGlnAspAlaGluTyValValPheAsnGlySerValGlyAlaLeuThr 256
| | | | | : : : | | | | | : : : | | | | |
Db 713 GCCAAGCTGAAGAGAGCTCGACTACTACAACTTCAC----- 751
QY 257 GlyGluAsnAlaLeuGlyAlaValGlyGluThrValArgLeuPheValGlyAsnGly 276
| | | | | : : : | | | | | : : : | | | | |
Db 752 -----AAGCGCACCGTCTCGGCGACTTACAGACGCTCAGCGCCACGCGC 796

RESULT 26

US-09-328-352-1877
; Sequence 1877, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: GARY L. BRETON et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNI FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1877
; LENGTH: 2016
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1877

Alignment Scores:
Pred. No.: 0.000454 Length: 2016
Score: 132.00 Matches: 96
Percent Similarity: 36.09% Conservative: 74
Best Local Similarity: 20.38% Mismatches: 173
Query Match: 5.03% Indels: 128
DB: 4 Gaps: 23

US-10-088-045-2 (1-502) x US-09-328-352-1877 (1-2016)

QY 107 ThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlyAspThrIleGlu 126
| | | | | : : : | | | | | : : : | | | | |
Db 223 ACCGTAATGGAATAATTACAGACCACTCTTGAAGGTTGAAGAGGGGATGAGGCTGTT 282
QY 127 ValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPheHisAlaIle 146
| | | | | : : : | | | | | : : : | | | | |
Db 283 ATTCAGCTCATATATCAATTAAGATCA---GATACCTCTTTACATCGAGGCTTG 339
QY 147 ThrGlyProGly-----GlyGlyAlaGluAlaSerPheThrAlaPro 160
| | | | | : : : | | | | | : : : | | | | |
Db 340 TTACTGACAGGTTAATGATGATGCTGCGGGTTTAAATTAAAGAAATCGCGCA 399
QY 161 GlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyValTyHisCys 180
| | | | | : : : | | | | | : : : | | | | |
Db 400 AATGAGATTTTTAAACCGCTTAAAGTGAACAAATGCAATATGATGATGATC--- 456
QY 181 AlaValAlaProValGlyMetHisIleAlaSerGlyMetTyGlyLeuIleValGlu 200
| | | | | : : : | | | | | : : : | | | | |
Db 457 -----GCCATATGTAAGGCAAGACCAAGACGAGTTATATGAGCGCGCTTGATTTAT 510
QY 201 ProLysGluGlyLeuPro-----LysValAspLysGluTyTyValMet 215

Db 511 CCAAAAGGCAAAATACAGTTGGCGCGCATGAAAAAATCTATGCTATTAATGATGATG 570
QY 216 GlnGlyAspPheTyThrTyrGlyTyGlyGlyGlnGlyLeuGlnProPheAspMet 235
| | | | | : : : | | | | | : : : | | | | |
Db 571 CTTTCAGATTTTATATATCTTCTAGC-----GATAGTATT 606
QY 236 GlnLysAlaIleArgGlnAspAlaGluTyValPheAsnGlySerValGlyAlaLeuThr 245
| | | | | : : : | | | | | : : : | | | | |
Db 607 ATGAGAAATTGAAGAAATCTGCTGATATTATCAAAACCGCGTGAACCGTATGAT 666
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyGlu----- 258
| | | | | : : : | | | | | : : : | | | | |
Db 667 GTTTTAAGCAAGCTTAAGACAGCGGCTTAAAGCACTTGCAAGTCTTGATGCTG 726
QY 259 -----AsnAlaLeuTyGlyAlaValGlyGlu---ThrValArgLeuPheValGly 274
| | | | | : : : | | | | | : : : | | | | |
Db 727 AATCAGATGCGGATGTTAAAAACCATATGCTGATGTCAGGCTTATACCTTTTATGTT 786
QY 275 AsnGlyGly-----ProAsnLeuThrSerSerPheHisValIleGlyGluIlePhe 291
| | | | | : : : | | | | | : : : | | | | |
Db 787 AACGCAACACACCTCAGCAAAATCTGACAGATCTTTAAAGCTGCTGATAGTT--- 843
QY 292 AspLysValHisPheGluGlyGlyGlyGlyGlnHis-----Asn 305
| | | | | : : : | | | | | : : : | | | | |
Db 844 ---CGATTGCGATTTTATTAATGCATCAGCAATGCTATTTTGTATGATCAGAAATCCAAAT 900
QY 306 IleGlnThrThrLeuIleProAlaGlyGly-----AlaAlaIleThrGlu 320
| | | | | : : : | | | | | : : : | | | | |
Db 901 TTAATAATGCTGCTGTAAGTGAAGTGAAGGACAGCCAGCTTAACCTGTAGCAATGATGAG 960
QY 321 PheLysValAspValProGlyAspTyValLeuValAspHisAlaIlePheArgAlaPhe 340
| | | | | : : : | | | | | : : : | | | | |
Db 961 TTTGATATGCTACACTGAACCTTACGATGCTATT----- 996
QY 341 AsnLysGlyAlaLeuGlyIleLeuLysValGlyGlyGln---AsnHisGluIleTy 359
| | | | | : : : | | | | | : : : | | | | |
Db 997 -----GTGCAACCTTAACAGCCAAATTCAGATT--- 1026
QY 360 SerHisLysGlnThrAspAlaValTyLeuProGlnGlyValaProGlnAlaIleAspThr 379
| | | | | : : : | | | | | : : : | | | | |
Db 1027 -----GACCAAGCTCTATT-----GATGTAAGTGTGTTGCTATCGGCTACT 1068
QY 380 GlnGluAlaProLysProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLys 399
| | | | | : : : | | | | | : : : | | | | |
Db 1069 TTGCACAATGAATAATACGACAGCTGTGGCGCCAGTCCAAATGCCATACCCGCCACT 1128
QY 400 Ala-----ThrTyAsp-----SerAsnCysAlaAla 408
| | | | | : : : | | | | | : : : | | | | |
Db 1129 GCTTATATGACCATGATGATATGGCATGATGATGATGATGATGATGATGATGATGAT 1188
QY 409 CysHisGlnProAspGlyLysGlyValProAsnAlaPheProProLeu----- 424
| | | | | : : : | | | | | : : : | | | | |
Db 1189 GGCATCAATGAATGAACATGACATGATGCTGCGCCATGCTGAATGAAGAAAGAGAG 1248
QY 425 -----AlaAsnSerAspTy-----LeuAsn 431
| | | | | : : : | | | | | : : : | | | | |
Db 1249 ACTCAGGCTATATATATATACACAAATGATGATGATGATGATGATGATGATGATGAT 1308
QY 432 AlaAsp---HisAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLysIleThr 450
| | | | | : : : | | | | | : : : | | | | |
Db 1309 TCTGACGGCATGATCATTCATGATGACATGACATGACATGATGATGATGATGATGATGAT 1368
QY 451 ValAsnGlyAsnGlnTyGlyLysValMetProAlaIleAlaLeuSerAspGlnGlnIle 470
| | | | | : : : | | | | | : : : | | | | |
Db 1369 ATTAACCATGAT-----ATGCAAGCAATGATGATGATGATGATGATGATGATGAT 1413
QY 471 AlaAsnValIleThrTyThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAla 490
| | | | | : : : | | | | | : : : | | | | |
Db 1414 GCTATGATG-----AATATGATGATGATGATGATGATGATGATGATGATGATGAT 1446
QY 491 AspAspValAlaLysAlaLysTyThrLysPro 501

Db 1447 CAATCTGAGAAATAGCTTAAAGAGATGACCA 1479

RESULT 27
US-09-328-352-3903
Sequence 3903, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER.
FILE REFERENCE: CTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 3903
LENGTH: 1908
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-3903

Alignment Scores:
Pred. No.: 0.000471 Length: 1908
Score: 131.50 Matches: 73
Percent Similarity: 35.36% Conservative: 49
Best Local Similarity: 21.16% Mismatches: 116
Query Match: 5.01% Indels: 107
Gaps: 19

US-10-088-045-2 (1-502) x US-09-328-352-3903 (1-1908)

Qy 107 ThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGlyAspThrIleGlu 126
Db 142 ACTGTAAATGTAAGTTGTTGACCTCTTCTGATGTAAGAGGATGATGATCCGTT 201
Qy 127 ValGlnPheSerArg-----HisProAspSerIleMetProHisAsnValAspPhe 143
Db 202 ATTCGCGTACACCAACAGTGTAAACCAACGATTTCA-----TCTATTCATGCG 249
Qy 144 HisAlaIaIaThrGlyProGly-----GlyGlyAlaGluAlaSerPhe 157
Db 250 CATGCTTAATTTATTAACCTGCAATTAATGATGATGATGATGATGATGATGATGAT 309
Qy 158 ThrAlaProGlyHisThrSerThrPheSerPheIleValLeuGlnProGlyLeuVal 177
Db 310 ATTCACCAATTAACCTGATGATTAATTAATTAAGGTTGTCAGAAATGAACTATTGG 369
Qy 178 TyrHisCysAlaValAlaProValGlyMetHisIle-----AlaAsnGlyMet 193
Db 370 TATACAC-----TCGCATGTAAGAGCCCAAGCAAGATGCGCTTG 408
Qy 194 TyrGlyLeuIleLeuValGluProGlyGlyLeuPro-----LysVal 208
Db 409 TATGCTCTAGTATTATTATCCCAAAATAAAGTTACTTTAAGCGCTGAGAAAAGCT 468
Qy 209 AspIleValGlyTyrValMetGlnGlyAspPheTyr-----ThrIleGlyLys----- 224
Db 469 GATAGAGATTCGAGTCTTACTCTGATTTTCAACATTTACTAGTGGTCAATATTATG 528
Qy 225 -----TyrGlyGlnGlnGlyLeuGlnProPheAspMet 235
Db 529 AGTAATCTCAAGAAAGAGCTGATTATTATCCAGATCGAAGAGAAAGCGTTATGAGT 588
Qy 236 GluIleValAlaIleArgGlyAsp----- 242
Db 589 TTCAAGCAGATTAAAGAGATGTTTAAAGCAACATGAAAAGATCGTTCCATGATGAT 648
Qy 243 -----AlaGlyTyrVal-----ValPheAsn 249
Db 649 CAGATGCGAATGCTTAAGACAGATATGTCGATGAACAACATTAAGTTTAAATGAT 708
Qy 250 GlySerValGlyAlaLeuThrGlyGluAsnAlaLeuIleValIleValIleVal 269
Db 709 GGTAAAGATCCGAAACAAATTTGAGCTGAAATTTTAAAGCT-----GGAGAAAAGTA 762

Qy 270 ArgLeu-----PheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGly 288
Db 763 GCTTCCTGTTTATTCATGATGTTCCGCAATGCTTTG----- 798
Qy 289 GluIlePheAspIleValHisPheGlyGlyLysGlyGluAsnHisValIleGlnThr 308
Db 799 -----TTCGATGTA-----AGAATTCCTAATCGAAAATG 828
Qy 309 ThrLeuIleProAlaGly-----AlaAlaIleThrGluPheLysVal 323
Db 829 ACAATGTCAGCCGCGATGACCAACAGTAAACCCGTCCAGTGAATGTCGATG 888
Qy 324 AspValProGlyAspTyr-----ValLeuValAsp-----HisAlaIlePheArg----- 338
Db 889 GGAACCGCTGACATTAATGATGATTTATTTGAAACCGAAACAAACAACTTATCAATGAA 948
Qy 339 -----AlaPheAsnIleGlyAlaLeuGlyTyrIleLeuIleValGluGlyGluAsnHis 356
Db 949 GCTGAATCTATTGATCGAAGTGTCTTCTATCGTACATTAATGATGATGAAATACTTCA 1008
Qy 357 GluIleTyrSerHisIleValThrAspAlaValTyrLeuProGlyAlaProGlnAla 376
Db 1009 CTTGTAATAAAT-----ATATGATGCGACACCTCGTCCCGTGT 1050
Qy 377 IleAspThrGlnGlu 381
Db 1051 TTACTGACGATGGA 1065

RESULT 28
US-09-252-991A-14169/c
Sequence 14169, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14169
LENGTH: 1323
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14169

Alignment Scores:
Pred. No.: 0.000339 Length: 1323
Score: 130.50 Matches: 82
Percent Similarity: 35.51% Conservative: 70
Best Local Similarity: 19.16% Mismatches: 152
Query Match: 4.97% Indels: 124
Gaps: 19

US-10-088-045-2 (1-502) x US-09-252-991A-14169 (1-1323)

Qy 117 IleArgValArgGlyGlyAspThrIleGluVal-----GlnPheSerAsnHis 132
Db 1108 GTTCGCGCGCGGCTCCGATTCAGTCCGCTTCCGACACATCTAGCGCAACCAATCAC 1049
Qy 133 -ProAspSerIleMetProHisAsnVal-----AspPheHisAlaAla 146
Db 1048 CCCGACCAAGAA-----CACGCAATCGAGCGCTACAGCCGCAAGAGATTTCCGCGC 995
Qy 146 AthrGlyProGlyGlyAlaGluAlaSerPheThrAlaProGlyHisThrSerThrPhe 166
Db 994 TCTTACGAGGCGCAAGGAGGACGCGCTTACTATTCGAGCATGCCCTTACACC-- 937

DB 1443 GTGAACAGCCGCTTCATCCCGGCTTACGCTGATCCCGAAGATGACTACACTGTC 1502
QY 215 MetcGlnGlyAspPheTyrThrGlyGlySerGlyGlyGlnGlyLeuGlnProPhe--- 233
DB 1503 CTCATCGGTCAGTGGTACCTAAG-----AGCCACACCCAGTTGAAGAGTTCCTC 1553
QY 234 AspMetGlnGlyValAlaIle---ArgGlnAspAlaGlnTyrValValPheAsnGlySerVal 252
DB 1554 GACGCTGCTCTACTATTGCTGCTCCAGACCGT-----ATTGTCATCAACGGAAGTCC 1607
QY 253 GAlaAlaLeuThrGlyGlnAsnAla-----LeuGlyAlaValGlyGlnThrValArg 270
DB 1608 GGAAGAGGTGATGATCAGACGACCCGCTTTCACCTTGAAGCTGGAAAGACTTACAGG 1667
QY 271 LeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlnIle 290
DB 1668 GTTAGATCTGTAACTGGGGGTGTCAGACATCTATCACTTACAGATTCAAGATCAGAG 1727
QY 291 PheAspGlyValHisPheGlnGlyGlyGlyGlnGlnHisAsnIleGlnThrThrLeu 310
DB 1728 ATGAAGCTGCTTGAATGGAAGATCGACCTTCTCAAAAGATACGACTCTTGAC 1787
QY 311 IleProAlaGly-----GlyAlaAlaIleThrGlnPheGlyValAspValProGly 327
DB 1788 GTTCACGTTGGCCAGTGGCTTGGACACCATCGTACCGGAGATCAA-----GAACCTAA 1841
QY 328 AspGlyValLeuValAspHisAlaIlePheArgAlaPheAsnGlyGlyAlaLeuGlyIle 347
DB 1842 GATTACATCAATGATGTTGATCTCTTGAAGGTTTATCAACAACCGGACTT 1901
QY 348 LeuGlyValGlnGly 352
DB 1902 CTCGCTACAGAGGA 1916

RESULT 30
US-08-746-111-4
Sequence 4, Application US/08746111
Patent No. 6066778
GENERAL INFORMATION:
APPLICANT: Ginsburg, David
TITLE OF INVENTION: Compositions And Methods For Screening
TITLE OF INVENTION: Compounds For Anticoagulant Activity
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medien & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,111
FILING DATE: 06-NOV-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: 04-02536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6585 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 6..6554
US-08-746-111-4
Alignment Scores:
Pred. No.: 0.00471 Length: 6585
Score: 130.00 Matches: 70
Percent Similarity: 34.32% Conservative: 46
Best Local Similarity: 20.71% Mismatches: 120
Query Match: 4.95% Indels: 102
Gaps: 16
US-10-088-045-2 (1-502) x US-08-746-111-4 (1-6585)
QY 63 AspAlaIleValThrHisAlaProGlnValProProValAspArg----- 78
DB 1074 GAGGTATTTGGAATGATGACCCGGATGCTCGCGAATATGACAAATTTACAGGTCT 1133
QY 79 -----AspHisProAlaGlyValValValys 87
DB 1134 CAGCAGCTGGATTAATTTCTCAACCAAAATTGGAACATTAACAAGAAAGTTATCTACAGG 1193
QY 88 MetGlnThrValGlnGlyValMetArgLeuAlaAspGlyValGlnTyrGlnPheThr 107
DB 1194 CAATATGAGAAAGAGCTTACACCAAGCATGACAAACCCAGCATCAA----- 1244
QY 108 PheGlyGlyGlnValProGlyGlnMetIleArgValArgGlnGlyAspThrIleGlnVal 127
DB 1245 ---CAAGTGGAGTTCTGGGCGCTGTTATCAGAGCCAGGTGACAGACACACTCAAGATC 1301
QY 128 GlnPheSerAsn-----HisProAspSerLysMetProHisAsnValAspPheHis 144
DB 1302 GTGTCAAAAATATGCGAGCCGACCTTACAGCATTTTACCTCAGCGGGTGAACCTTCTCT 1361
QY 145 AlaAlaThrGlyProGlyGlyGlyAlaGlnAlaSerPheThrAlaProGlyHisThrSer 164
DB 1362 CTTTACGAA-----GATGGAATCAATTTCTTCCACCTCAGAGCTCACACC--- 1409
QY 165 ThrPheSerPheLysAlaLeuGlnProGly-----LeuTyrValTyr----- 178
DB 1410 -----ACAGTACAGACAGATTCACCGGGGAACCTTCACTTCAAAATGAGACATTCTA 1463
QY 179 -----HisCysAlaValAlaPro----- 184
DB 1464 GAGTTTGAATGACCAACGGAACGATGCCAGTGGCTTACCAAGGCCACTACTACAGTAT 1523
QY 185 -----ValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuValGlnProLys 202
DB 1524 GTGAGCTTACAAAGGATATGCTCTGGGCTGATAGGGCTGCTTAAATTGTAAGAC 1583
QY 203 GlnGlyLeuProLys-----ValAspLysGlnTyrTyrValMetGln 216
DB 1584 AGTCTCTGACACGAGGGGTGATCAGAGGGTGGCAGACATGACGACGACGCGGTGTT 1643
QY 217 GlyAspPhe---TyrThrLysGlyLysTyrGlyGlnGlnGlyLeuGlnProPhe----- 233
DB 1644 GCTGTGTTGACGAGAAAGAGCTGTACATTGAGAGCAACATCAACAAAGTTCTGTGAG 1703
QY 234 AspMetGlnGlyValAlaIleArgGlnAspAlaGlnTyrValValPheAsnGlySerValGly 253
DB 1704 AATCTGTATGAGTGAAGCTGATGCCAAGTTT----- 1739
QY 254 AlaLeuThrGlyGlnHisAlaLeuLysValAlaValGlnThrValArgLeuPheVal 273
DB 1740 -----TACGAATCAAAATATGACACTTACACGCTTACAGTCTG 1778
QY 274 GlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGlnIlePheAsp--- 292
DB 1779 -----CCGAGAGCATTTCCACTGTGGATTTGTTGATGAC 1817

QY 386 ProAlaProAla-----AsnLeuGlnGlnGlnIleValGlyLys 399
Db 2194 ACAGTCATGGCTACACGGAATAATGATGATGTTAGAACTGAAGATGAT 2253
QY 400 AlaThrTyrAsp 403
Db 2254 GCTGACTATGAT 2265

RESULT 32
US-08-658-340-1
Sequence 1, Application US/08658340
Patent No. 5910576
Patent No. 5910576 5861489
GENERAL INFORMATION:
APPLICANT: Bertina, Rogier
APPLICANT: Reitsma, Pieter
TITLE OF INVENTION: A method for diagnosing an increased
TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
TITLE OF INVENTION: thrombosis and kit for use with the same.
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Akzo No. 5910576 5861489el Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,340
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/454,353
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gornley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 948-7400
TELEFAX: (301) 948-9751
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6909 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-658-340-1

Alignment Scores:
Pred. No.: 0.00732 Length: 6909
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55
Best Local Similarity: 20.05% Mismatches: 145
Query Match: 4.89% Indels: 139
Gaps: 21
US-10-088-045-2 (1-502) x US-08-658-340-1 (1-6909)

QY 63 AspAlaIleValThrHisAlaProGluValProProProValAspArg-----78
Db 1162 GAAGTCATTTGGAGTACGACCTGTAATACACGCAATATGACAAAAATACAGCT 1221
QY 79 -----AspHisProAlaLysValValValLys 87
Db 1222 CAGCATTTGATTAATTTCTCAACCAATTGGAAAAACATTATTAAGAAAGTTATGTACA 1281
QY 88 Met-----GlnThrValGlnLysValMetArgLeuAlaAsp 99

Db 1282 CAGTACGAGATGATGCTCTTACCAACAATACAGTAAATCCAAATATGAAA---GAAGAT 1338
QY 100 GLYValGlnTyrGlnThrThrPheThrPheGlnGlnValProGlnIleMetIleArgVal 119
Db 1339 GGG-----ATTGGGCTCTTATATCAAGACC 1365
QY 120 ArgGlnGlyAspThrIleGlnValGlnPheSerAsn-----HisProAspSerLys 136
Db 1366 CAGTCAGACACACACTCAAAATCGCTTCAAAAATATGCGCCGCCCTTATACATT 1425
QY 137 MetProHisAsnValAspPheHisAlaIleThrGlyProGlnGlyGlnAlaGlnLys 156
Db 1426 TACCTCATGAGATGACTCTTCTGCTTATGAAGATGA-----GTCACTCTTCT 1476
QY 157 PheThrAlaProGlnIleThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu 175
Db 1477 TTGACCTCA---GGCAGGAACAACACCATG---ATCAGAGCAGTTCAACCGGGAACC 1530
QY 176 TyrValTyr-----HisCys 180
Db 1531 TATACCTTATAGTGAACATCTTAGAGTTGATGAACCCACAGAAAATGATGCCAGTGC 1590
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
Db 1591 TTAACAAGACCACTACTACAGTACAGTACATGATGATGAGACATCCCTCTGGCTAAT 1650
QY 195 GlnLeuIleLeuValGlnProLysGlnGlyLeuProLys-----Val 208
Db 1651 GAGTACTCTTAATCTGTAAGACAGATCCCTGACAGCGGAGAAATACAGAGCGACA 1710
QY 209 AspLysGlnTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGln 227
Db 1711 GACATGGAACAGAGCGCTGCTTGTGCTTTGATGAAGAACAAAGCTGTCCTTGAG 1770
QY 228 GlnGlyLeuGlnProPhe---AspMetGlnLysAlaIleArgGlnAspAlaGlnTyr 245
Db 1771 GACACATCAACAAGTTTGTGAAATCCGATGAGATGAGAAAGTATGACCCCAAGTTT 1830
QY 246 ValValPheAsnGlySerValGlnAlaLeuThrGlnAsnAlaLeuLysAlaVal 265
Db 1831 -----TATGAATCAAAATC 1845
QY 266 GlyLysThrValArgLeuPheValGlnAsnGlyLysProAsnLeuThrSerPheHis 285
Db 1846 ATGAGCACTATCAATGCTATGTC-----CCTAGACATTAATCT 1884
QY 286 ValIleGlyLysIlePheAsp-----292
Db 1885 ACTCTTGATTCGCTTGTATGACACTGTCAGTGCACCTTGTATGTCGGGACCCAG 1944
QY 293 -----LysValHisPheGlnGlyLys-----GlyGlnAsnHisAsn 305
Db 1945 AATGAATTTTGAACATCACTTCACTGGGCACTTCACTATGAGAAAGAGCAGTAC 2004
QY 306 IleGlnThrThrLeuLeuPheProAlaGlyGlnAlaAlaIleThrGlnPheLysValAspVal 325
Db 2005 GACACCTTGACCTCTTCCCAATGCGTGAAGATCTGTACG-----2046
QY 326 ProGlnAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlnAlaLeu 345
Db 2047 -----GTCACAATGAT-----AATGTTGAACCTTGG 2073
QY 346 GlyIleLeuLysValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 365
Db 2074 ATGTAACTTCCATGATTTCTAGTCCAAAGAAAGAGAGAGAGAGAGAGAGAGAGAT 2133
QY 366 AlaValTyrLeuProGlnGlnGlnAlaProGlnAlaIleAspThrGlnGlnAlaProLysThr 385
Db 2134 GTTAAATGATATCCCAATGATGATGAAGACTCATAGAGATTTTGAACCTCCAGAAATCT 2193
QY 386 ProAlaProAla-----AsnLeuGlnGlnGlnIleValGlyLys 399

Db 2194 ACAGTATGCTACACGAAATGATGATGTTAGAACCTGAAGATGAGAGTGTAT 2253
 QY 400 AlAtHrTYrAsp 403
 Db 2254 GCTGACTATGAT 2265

RESULT 33

US-08-746-111-26
 ; Sequence 26, Application US/08746111
 ; Patent No. 6066778
 ; GENERAL INFORMATION:
 ; APPLICANT: Ginsburg, David
 ; APPLICANT: Cui, Jisong
 ; TITLE OF INVENTION: Compositions And Methods For Screening
 ; TITLE OF INVENTION: Compounds For Anticoagulant Activity
 ; NUMBER OF SEQUENCES: 54
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Medlen & Carroll, LLP
 ; STREET: 220 Montgomery Street, Suite 2200
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/746,111
 ; FILING DATE: 06-NOV-1996
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ingolia, Diane E.
 ; REGISTRATION NUMBER: 40,027
 ; REFERENCE/DOCKET NUMBER: UM-02536
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ. ID NO: 26:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 6909 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-746-111-26

Alignment Scores:

Pred. No.: 0.00732 Length: 6909
 Score: 128.50 Matches: 85
 Percent Similarity: 33.02% Conservative: 55
 Best Local Similarity: 20.05% Mismatches: 145
 Query Match: 4.89% Indels: 139
 DB: 3 Gaps: 21

US-10-088-045-2 (1-502) x US-08-746-111-26 (1-6909)

QY 63 AspAlaIleValThrHisAlaProGluValProProProValAspArg----- 78
 Db 1162 GAAGTCATTGGAGCTATGACCTGTATATCCAGCAATATGCAAAATAACAGTCT 1221
 QY 79 -----AspHisProAlaIleValValValValValValValValValVal 87
 Db 1222 CAGCATTTGATTAATTTCTCAAAACAATTGGAAAACATTATTAAGAAAGTTATGACACA 1281
 QY 88 Met-----GluThrValGluIleValMetArgIleValAsp 99
 Db 1282 CAGTACGAAGATGAGCTCTTACCAACAATACAGTAATCCCAATATGAAA---GAAGAT 1338
 QY 100 GlyValGluIleValPheThrPheGlyGlyGlyValProGlyGlnMetIleArgVal 119
 Db 1339 GGG-----ATTTGGGTCTCTATATCAAGGCC 1365

QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsp-----HisProAspSerLys 136
 Db 1366 CAGGTACAGACACACTCAAAATCGCTTCAAAAATATGCGCCGCCCTTATACATT 1425
 QY 137 MetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValGluValAsp 156
 Db 1426 TACCTTCAAGATGAGTGCCTTCTCGCTTATGAAGATGAA-----GTCAACTCTTCT 1476
 QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheValAlaGlnProGly---Leu 175
 Db 1477 TTCACCTCA---GGCAGAAACAACACCATG---ATCAGACAGTTCAACCGGGAAC 1530
 QY 176 TyrValTyr-----HisCys 180
 Db 1531 TATACTTAAATGAGAACATCTTAGAGTTGATGAACCAACAGAAATATGCCAGTGC 1590
 QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
 Db 1591 TTACACAGACCACTACTACAGTACGAGTGCATGATGAGACATTCGCTCGGCTTAATA 1650
 QY 195 GlyIleIleLeuValGluProIleGlyLeuProLys-----Val 208
 Db 1651 GGACTACTTATATCTGTAAGACAGATCCCTGCACAGCGGAGAAATACAGAGGAGCA 1710
 QY 209 AspLysGluTyrTyrValMetGlnIleAspPhe---TyrThrLysGlyLysTyrGly 227
 Db 1711 GACATGGAACAGACAGCTGTGTTGCTGCTGTGATGAACAACAAAGCTGTCCTTGA 1770
 QY 228 GlnGlyLeuGlnProPhe-----AspMetGluLysAlaIleArgGluAspAlaGly 245
 Db 1771 GACAACATCAACAGATTGTGAAATCCTGATGAGGTGAACCGTATGATGCCCAAGTT 1830
 QY 246 ValValPheAsnGlySerValGlyValAlaLeuThrGlyGluAsnAlaLeuLysVal 265
 Db 1831 -----TATAAATCAAAATC 1845
 QY 266 GlyIleThrValArgLeuPheValGlyAsnGlyLysProAsnLeuThrSerSerPheHis 285
 Db 1846 ATGAGCAGCTATCAATGAGCTATGTG-----CCTGAGACATTAAT 1884
 QY 286 ValIleGlyIlePheAsp----- 292
 Db 1885 ACTCTTGATTCGCTTATGACACTGTCAGTGCAGTCTCTAGTGTGGGACCCAG 1944
 QY 293 -----LysValHisPheGluGlyLys-----GlyGluAsnHisAsn 305
 Db 1945 AATGAATTTTGACCATCCACTTCACTGGGCACTCATCTATGATGAAGAGGCATGAG 2004
 QY 306 IleGlnThrThrLeuIleProAlaGlyAlaIleThrGluPheLysValAspVal 325
 Db 2005 GACACCTTGACCTCTTCCCACTGCGTGAAGATCTGTACG----- 2046
 QY 326 ProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345
 Db 2047 -----GTACAAATGAT-----AATCTGGAACCTGG 2073
 QY 346 GlyIleLeuLysValGluGlyGluGluAsnHisGlyIleTyrSerHisLysGlnThrAsp 365
 Db 2074 ATGTTAACTTCATGAATTTCTAGTCCAGAAACAAGCTAGAGCTGAATTCAGAGGAT 2133
 QY 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThr 385
 Db 2134 GTTAAATGATCCCAAGATGATGAAGAATCATATGAGATTTTGAACCTCCAGAAATGT 2193
 QY 386 ProAlaProAla-----AsnLeuGlnGluGlnIleLysAlaGlyLys 399
 Db 2194 ACAGTCATGCGTACACGAAATGATGATGCTTTAGAACTGAAGATGAAGAAGTGTAT 2253
 QY 400 AlAtHrTYrAsp 403
 Db 2254 GCTGACTATGAT 2265

```

US-08-454-353A-1
/ Sequence 1, Application US/08454353A
/ Patent No. 6518016
/ GENERAL INFORMATION:
/ APPLICANT: Bertina, Rogier
/ APPLICANT: Bertina, Pieter
/ TITLE OF INVENTION: A method for diagnosing an increased
/ TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
/ TITLE OF INVENTION: thrombosis and kit for use with the same.
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Akzo No. 6518016el Patent Department
/ STREET: 1300 Piccard Drive, Suite 206
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/454,353A
/ FILING DATE: 06-JUN-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Gormley, Mary E.
/ REGISTRATION NUMBER: 34,409
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 948-7400
/ TELEFAX: (301) 948-9751
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 6909 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: unknown
/ MOLECULE TYPE: cDNA
/ US-08-454-353A-1
Alignment Scores:
Pred. No.: 0.00732 Length: 6909
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55
Best Local Similarity: 20.05% Mismatches: 145
Query Match: 4.89% Indels: 139
DB: Gaps: 21
US-10-088-045-2 (1-502) x US-08-454-353A-1 (1-6909)
QY 63 AspaLaIleValThRHisAlaPProGluValProProValAspArg----- 78
::: ||| ::||| ::||| ::||| ::|||
Db 1162 GAAGCATTTGGAGACTGATGCACTGTAATACGAGGAATATGCAAAAAAATACAGCTT 1221
79 -----AspHisProAlaIleValValVallys 87
1222 CAGCATTTGGATATATTCTCAAAACCAAAATTGGAAAAACATTAAGAAAATTATGACACA 1281
QY 88 Met-----GluThrValGluValMetArgLeuLaaSer 99
Db 1282 CAGTACGAAGATGAGTCTTCAACCAACATACAGTAATCCCAATATGAAA--GAAAT 1338
QY 100 GtValGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1339 GCG-----ATTGGTGCCTATATATGAGCC 1365
QY 120 ArgGluGtAspThrTleGluValGlnPheSerAan-----HisProAspSerLys 136
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1366 CAGGTACAGACACACTCAAAATCGTGTCAAAAATATATGCGACGCCGCCCTATGCAATT 1425
137 MetProHisAenValAspPheHisAlaIleThrGtGtProGtGtGtGtGtGtGtGtGtGt 156
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
::: |||

```

Db 1426 TAACCTCATGGAGGACCTTCTGCGCTTATGAAGATGA-----GTCACCTCTCT 1476
 Oy 157 PheThx1aLapProGluYhiStrSerThrPheSerPheYsa1aLeuGlnProGly--Leu 175
 Db 1477 TTCACCTCTCA---GGCAGGAACAACACCATG--ATCAGAGCAGTTCAACAGGGGAACC 1530
 Oy 176 TyrValYyr-----HisCys 180
 Db 1531 TATCTCTTATAGTGGAAACATCTTAGAGTTGATGAACCCACAGAAAATGATGCCCACTGC 1590
 Oy 181 AlaVala1aLapPro-----ValGlywchis1lealaenGlymetYr 194
 Db 1591 TTAAACAAGACCATCTACAGTGAAGCTGGACATCATGAGACATCGCTCGGGCTATATA 1650
 Oy 195 GlyLeu1leuValaGluProLySGluGlyLeuProLyS-----Val 208
 Db 1651 GGACTCTCTTAACTCTTAATCTTGAAGAGCAGATCCCTGGACGGAGATACAGAGGCGACA 1710
 Oy 209 AspLySGluYrYrYrValMetGlnGlyAspPhe---TyrThrLySGlyYrGlyGlu 227
 Db 1711 GACATCGCAACAGCAGCAGCTGTGTGTTGCTGTGTGTTGATGAACAAAGCTGTACTTGAG 1770
 Oy 228 GlnGlyLeuGlnProPhe-----AspMetGlyuLySa1a1ealArguAspAlaGluYr 245
 Db 1771 GACAACATCAACAAGTTTGTGAAAAATCGATGAGGTGAAGAGTATGACCCCAAGATT 1830
 Oy 246 ValValaPheasnGlySerValGlyAlaLeuThGlyGluasnAlaLeuLySa1aLySa1 265
 Db 1831 -----TATGAATCAACATC 1845
 Oy 266 GlyYrThrYrValArgLeuPheValGlyasnGlyYrProasnLeuThrSerSerPheHis 285
 Db 1846 ATGAGCAGCTATCAATGAGCTATGTC-----CCTGAGAGCATTAAT 1884
 Oy 286 Val1leGlyGlu1lePheAsp----- 292
 Db 1885 ACTCTTGATCTTCGCTTGTATGACACCTGCCAGTGGCAGCTTCGTAGTGGGAGCCAG 1944
 Oy 293 -----LySa1a1ePheGluGlyLyS-----GlyGluasnHisAsn 305
 Db 1945 AATGAATTTTGAACATCACTCACTTCACTGGGCACATCTATGGAAGAAGGCATGAG 2004
 Oy 306 IleGlnThrThrLeu1leProAlaGlyGlyAla1a1ePheGluPheLySa1aAspVal 325
 Db 2005 GACACCTTGACCCCTTCCCATGCGTGGAGAACTGTGACG----- 2046
 Oy 326 ProGlyAspYrYrValLeuValaPheHisAla1lePheArgAlaPheasnLySGlyAlaLeu 345
 Db 2047 -----GTCACAAATGAT-----AATGTTGGAACCTGG 2073
 Oy 346 Gly1leuLeuYsa1aGluGlyGluGluasnHisGlu1leYrSerHisLySGlnThrAsp 365
 Db 2074 ATGTTAACTTCCATGATTTCTAGTCCAAAGACAAAGAAGCTGAGAGCTGAATTCAGGGAT 2133
 Oy 366 AlaValaYrLeuPProGluGlyAlaProGlnAla1leAspThrGlnGlyu1a1aProLySThr 385
 Db 2134 GTTAAATGTATCCCAATGATGATGAAGAGCTCATATGAGATTTTGAACCTCCAGAAATCT 2193
 Oy 386 ProAlaProAla-----AsnLeuGlnGluGln1leLySa1aGlyLyS 399
 Db 2194 ACACTCATGGCTACAGAGAAAATGATATCGTTTAGAACCTGAAGATGAAGAAGTGTAT 2253
 Oy 400 AlaThrYrAsp 403
 Db 2254 GCTGACTATGAT 2265
 RESULT 35
 US-09-165-019-1
 ; Sequence 1, Application US/09165019
 ; Patent No. 6558913
 ; GENERAL INFORMATION:
 ; APPLICANT: Bertina, Rogier
 ; APPLICANT: Reitsma, Pieter
 ;


```

: TITLE OF INVENTION: A method for diagnosing an increased
: TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
: NUMBER OF INVENTION: thrombosis and kit for use with the same.
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/165,019
: FILING DATE:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/454,353
: FILING DATE:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6909 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: US-09-165-019-1

```

```

Alignment Scores:
Pred. No.: 0.00732 Length: 6909
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55
Best Local Similarity: 20.05% Mismatches: 145
Query Match: 4.89% Indels: 139
DB: 4 Gaps: 21
US-10-088-045-2 (1-502) x US-09-165-019-1 (1-6909)

```

```

QY 63 AspalallevaIhrihsalaprogluvalProProbovalasparg----- 78
DB 1162 GAAGCATTTGGAGCATATGACCTGTAATACAGCGAATATGACAAAATAACAGGCTT 1221
QY 79 -----AsphsProalalyvalvallye 87
DB 1222 CAGCATTTGGATATTTCTCAACCAATTGGAATAATTAAGAAAGTTATGTACACA 1281
QY 88 Met-----GluThryValGluysValMetArgLeuAlaAsp 99
DB 1282 CAGTAGAAGATGAGCTCTTACCAACATACAGTGAATCCCATATGAAA-----GAAGAT 1338
QY 100 GlyValGluThryGlnPheThryPheGlyGlyGlnValProGlyGlnMetLeuVal 119
DB 1339 GGG-----ATTTGGGTCTCTATTATCAGAGCC 1365
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsp-----HisProAspSerLys 136
DB 1366 CAGGTACAGACACACTCAAAATCGTTCAAAAATATGGCCAGCCGCTTATAGCATT 1425
QY 137 MetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyAlaGluAlaSer 156
DB 1426 TACCTCATGAGAGACCTTCTGCGCTTATGAGATGAA-----GTCACTCTCT 1476
QY 157 PheThralaProGlyHisThrSerThyPheSerPheLysAlaLeuGlnProGly--Leu 175
DB 1477 TTCACCTCA---GGCAGGAACAACACCATG---ATCAGACAGATTCAACAGGGGAAC 1530
QY 176 TyrValTyr-----HisCys 180
DB 1531 TATACTTATAGTGAACATCTTAGAGTTGATGAACCCACAGAAAATGATGCCAGTGC 1590
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetLys 194
DB 1591 TTAACAAGACCATCTACAGTGAAGCTGTGACATCATGAGAGACATCGCTCTGGCTATA 1650
QY 195 GlyLeuIleLeuValGluProLysGlyGlyLeuProLys-----Val 208
DB 1651 GGACTACTTCTTAATCTGTAAAGACGAGATCCCTGACAGCGGAGGAATACAGAGGCAGCA 1710

```

```

QY 209 AspLysGluThryTyrValMetGlnLysAspPhe---TyrThrLysGlyLysTyrGlyGlu 227
DB 1711 GACATGAAACAGACAGCTGTGCTGTGTTGATGAGAACAAAAGCTGTACCTTGAG 1770
QY 228 GlnGlyLeuGlnProPhe-----AspMetGluLysAlaIleArgGluAspAlaGlyTyr 245
DB 1771 GACAAATCAACAACTTTGTGAAAATCTGTGAGAGGTGAACGTGATGACCCCAAGTTT 1830
QY 246 ValValPheAsnLysSerValGlyAlaLeuThrGlyGluAsnAlaLeuLysAlaVal 265
DB 1831 -----TATGATCAAAATC 1845
QY 266 GlyGluThryValArgLeuPheValGlyAsnGlyLysProAsnLeuThrSerPheHis 285
DB 1846 ATGAGACATATCAATGAGCTATGAG-----CCTGAGACATATCT 1884
QY 286 ValIleGlyGluIlePheAsp----- 292
DB 1885 ACTCTTGATTTCTGCTTATGATGACACTGTCCAGTGCACCTTCTGATGTGGGAGCCAG 1944
QY 293 -----LysValHisPheGluGlyLys-----GlyGluAsnHisAsn 305
DB 1945 AATGAAATTTGACACATCACTTCACTGAGCACTATTCATGTGAAAGAGCAATGAG 2004
QY 306 IleGlnThrThryLeuIleProAlaGlyAlaAlaIleThrGluPheLysValAspVal 325
DB 2005 GACACCTTGACCTCTTCCCAATCGCTGAGATCTGTAGC----- 2046
QY 326 ProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345
DB 2047 -----GTCACAATGAGT-----AATGTTGGAACCTTG 2073
QY 346 GlyIleLeuLysValGluGlyGluGluAsnHisGluIleTyrSerHisLysGlnThrAsp 365
DB 2074 ATGTACTCTTCATGATATCTTACTGCAAGAGCAAAAGCTGAGCTGAATTCAGGAGAT 2133
QY 366 AlaValTyrLeuProGlyAlaProGlnAlaIleAspThrGlnGlnAlaProLysThr 385
DB 2134 GTTAAATGATATCCCAATGATGATGAAAGCTCATATGAGATTTTGAACCTCCAGATCT 2193
QY 386 ProAlaProAla-----AsnLeuGlnGluGlnIleLysAlaGlyLys 399
DB 2194 ACAGTCATGGCTTACAGGAATAATGATGATCGTTAGAACCTGAAGATGAAGAGATGAT 2253
QY 400 AlaThr-TyrAsp 403
DB 2254 GCTGACTATGAT 2265

```

```

RESULT 36
US-09-252-991A-6517
: Sequence 6517, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 6517
: LENGTH: 933
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-6517
Alignment Scores:
Pred. No.: 0.000663 Length: 933

```

Score:	125.50	Matches:	68
Percent Similarity:	33.98%	Conservative:	37
Best Local Similarity:	22.01%	Mismatches:	89
Query Match:	4.78%	Indels:	115
DB:	4	Gaps:	15

US-10-088-045-2 (1-502) X US-09-252-991A-6517 (1-933)

OY	187	MeChIStIAIAaAnsgIyMeTyrGlyLysIleValGluProLysGluGlyLeuPro	206
DB	313	ATGAAATGCGCCGAGAGAAAGTACGGGCGCATCTTGGC-----	351
OY	207	LysValAspLysGluTyrTyrValMetGlnLysPheTyrThrLysGlyLysTyrGly	226
DB	352	-----AAGTACGCGCGCATGCGCATGAGAAAGTGGCAAGAGACGAGCGCC	399
OY	227	GluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAlaGluTyrVal	246
DB	400	-----ATGAAATGGCGCGCGCATGTTCGGCCACTTGGC	435
OY	247	---ValPheAsnGlySer-----ValGlyAlaIleuThrGluLysAsnAlaLysVal	263
DB	436	TGGATCTGGCAGCGTTCGCCAGCGCAAGAGCGCCCTC-----	471
OY	264	LysValGlyLysThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSer	283
DB	472	-----GGTTCGCCAAGCTGGCGGCAAC	495
OY	284	PheHisValIleGlyGluIlePheAspLysValHisPheGluGlyLysGlyLysAsn	303
DB	496	-----GAGTGGCGCTGGGCG--GGTGACCGG	519
OY	304	HisAsnIleGlnThrThrLeuIleProAlaGlyValAlaAlaIleThrGluPheLysVal	323
DB	520	CAGAGATGAGACGCCCATCTTC-----GGCGGTGC-----	555
OY	324	AspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGly	343
DB	553	-----CACCGATCATGGCGGCTGGGCG-----	576
OY	344	AlaLeuGlyTyrLeuLysValGluGlyGluLysAsnHisGluIleTyrSerHisLysGln	363
DB	577	-----GACATCTCGCGCGAAGCGGGGTGAAACGTCGCGCTACGTCCGACCGAA	630
OY	364	ThrAspAlaValTyrTyrProGluGlyValAProGlnAlaIleAspThrGlnGluAlaPro	383
DB	631	CTCGCGCGGCGTGAAGTCCCGGAAGGACG-----	666
OY	384	LysThrProAlaProAlaAsnLeuGlnGluGlnIleLysAlaGlyLysAlaThrTyrAsp	403
DB	661	-----AAGCGGACGTGAGCGCGGCAAGCATGTTCTCG	696
OY	404	SerAsnGlyAlaAlaCysHisGlnProAspGlyLys-----GlyValPro	418
DB	697	GTCMACTGCGTCCGCTGCCACGCGCGCGAAGGCAAGGTACCGCGCTGGTGGCGCGCG	756
OY	419	AsnAlaPheProProLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAla	438
DB	757	AACCTGACCAACCTCC-----GGCGCTTATCTACGCGTCCAGCTACGCCAGCTCCAA	811
OY	439	SerIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSer	458
DB	811	CAAAACATCGCGCATGGCGCGCAAGGCGCAG-----	841
OY	459	ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThrTyr	476
DB	841	---ATGCGCGCGCGAGAGCATCTCGGGGCAAGAAAGGTGACATCTGGCGCGCTAC	897
OY	477	ThrLeuAsnSerPheGlyAsnLysGly	485
DB	898	ATCTACACCTGTGCAACACAGGCG	924

```

US-09-324-867-1
; Sequence 1, Application US/09324867A
; Patent No. 6251632
; GENERAL INFORMATION:
; APPLICANT: Lillcrap, David
; APPLICANT: Cameron, Cherie
; APPLICANT: No. 6251632ley, Colleen
; APPLICANT: Horrocks, L. Suzanne Hoyle
; APPLICANT: Hough, Christine
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
; FILE REFERENCE: 1669, 0010002/JAG/BJD
; CURRENT APPLICATION NUMBER: US/09/324,867A
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: 09/035,141
; EARLIER FILING DATE: 1998-03-059
; EARLIER APPLICATION NUMBER: 60/039,953
; EARLIER FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;
; LENGTH: 7032
; TYPE: DNA
; ORGANISM: Canis familiaris
; FEATURES:
; NAME/KEY: CDS
; LOCATION: 1...7029
;
US-09-324-867-1

```

Alignment Scores:	
Pred. No.:	0.0528
Score:	120.50
Percent Similarity:	35.14%
Best Local Similarity:	2.104
Query Match:	4.59%
DB:	3
Gaps:	25
Length:	7038
Matches:	97
Conservative:	65
Mismatches:	176
Indels:	123
Gaps:	25

US-10-088-045-2 (1-502) X US-09-324-867-1 (1-7032

QY	68	h1e1a1aProGluVal1ProProProValAspArgAspHis-----	80
Db	1222	TATGCTCCCTCAGGCCCCCAGCCCCCAATGAATGAAGTCAATAAAATCTGTATTGGAACAAAT	1281
QY	81	---ProAlaValValValValValValMetGluThrVal-----	GluLys 93
Db	1282	GGTCTCAGCGGAGTGGTGAAGAAGTACAAAATAAATCCGATTTGGCGATACACAGATGAG	1341
QY	94	ValMetArgLeuAlaSerGluValGluTyrGlnPheThrPheArgLysGluValPro	113
Db	1342	ACATTTAAGATCGCTGAAGCATTCAGTATGAATCAGAAATCCGGAGCCTTTATCTTAT	1401
QY	114	GlyGlnMetLeuArgValArgGluGlyAspThrLeuGluValGlnPheSerAsn-----	131
Db	1402	GGAGAGAGTT-----GGAGACACACCTCGCATTTATATTGAAGAACCAAGCC	1446
QY	132	---HisProAspSerLysMetProHisAsnValAspPhe-----	HisAlaAla 146
Db	1447	AGCGGCGCATTAATCAATCTTACCTTCATGGAGTCAATTATGTCACTCTCTGCACACAGG	1506
QY	147	ThrGlyProGluGlyGlyAlaGluAlaSerPhe-----	ThrAlaProGluHisPheSer 164
Db	1507	AGATTGCCAAAGAGTGTGAACAATTGAAGAATGTGCCAATTCGCCGGGA-----	GAG 1560
QY	165	ThrPheSerPheLys-----	AlaLeuGlnProGly----- 174
Db	1561	ATATTCAAGTATTAATGACAGCTGACCGTTGAAGATGACCAACTAATACAGATCTCGG	1620
QY	175	---LeuTyrValTyrHisCysAlaValAlaProValGluMetHisIleAlaAsnGlyMet	193
Db	1621	TGCCTGACCCGATTTACTCAAGCTTCATTAAATCTGAGAGAGATCTAGCTTCAGAGCTC	1680
QY	194	TyrGlyLeuIleLeuValGlnProLysGluGlyLeuProLys-----	207
Db	1681	ATTGGCCCTCTTTCATCTGCTTACAAAGATCTGTGAATCAACAGAGAAACCAATGATG	1740


```

Db      1120  CGTGGCGCTCGTCAAA-----ACCGTCCGGAT 1094
               |||
Qy      224  sTyrGlyGluGlnGlyLeuGlnProPheAspMetGluValAlaIleArgLysAspAlaG 244
               |||
Db      1093  GCGCGCGCGAGATGTCGCCAG-----GTCCAGCGGAGCGCT 1055
               |||
Qy      244  uTyrValAlaPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuValAla 264
               |||
Db      1054  CCAGCGCCATTTCGTGGCCAGCGGACCGGAGTGGTACCATCTCGAACACGAG 995
               |||
Qy      264  sValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerP 284
               |||
Db      994  GGTGGCTGAG---GTCAAGGGGTGTGCGCGGCGGATCGTATGATAGACGCGCGCT 938
               |||
Qy      284  e-----HisValIleGlyGluIlePheAspLysValHisPheGluGlyGly 300
               |||
Db      937  GGTGGCGCATCAGCGGATGACGAGCGGTGTGCGCGCATCGACGAGGT----- 883
               |||
Qy      300  sGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrG 320
               |||
Db      882  -----CGACCCAGCTCTGTCCTCCGCGGTGTGTCGAGACA 842
               |||
Qy      320  uPheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaP 340
               |||
Db      841  CGTGAAGATCAGCTCCAGTCCGCGCGTGTGAGAGTGGCGGACCATCTCTTGAACTCGT 782
               |||
Qy      340  eAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsnHisGluIleTyrSe 360
               |||
Db      781  TGACCTGGCGCGCGAGTGGCGGTGCGGGCGGAGAGAGC----- 726
               |||
Qy      360  rHisLysGlnThrAspAlaValTyrLeuProGluGlyAla----- 373
               |||
Db      735  -----GATGCTGTGTAGACCCAGTGTGCTCATCGCTTTCCAGCAGT 689
               |||
Qy      374  -----ProGlnAlaIleAspThrGlnGluAlaProLys---ThrProAlaProAl 389
               |||
Db      688  GCTTGTGTGTCGACGAGCCGCGTGTGTCGAGCAGCTCGACCTGACCCGAGCGCGG 629
               |||
Qy      389  aAsnLeuGlnGlnGlnIleLysAlaGlyLysAlaThrTyrAspSerSerGlyAlaIa 409
               |||
Db      628  GGATGTGCGCGCAGCGCTC-----GGCGT 605
               |||
Qy      409  sHisGlnProAspGlyLysGlyValProAsnAlaPheProPheLeuAlaAsnSerAspT 429
               |||
Db      604  TCATCAAGCC---GCCAAGTCCCGGAGACCGCTCCGGAGCTGTGATGGG----- 553
               |||
Qy      429  rLeuAsnAlaAspHisAlaArgAlaAlaSerIleValAlaAsnGlyLeuSerGlyLys 449
               |||
Db      552  -----CATGCTGAGCGCGCGCAGGTG-----GGCTTCGTAGATCAGC 515
               |||
Qy      449  eThrValAsnGlyAsnGlnTyrGluSerValMet 460
               |||
Db      514  TCGCGTCCAGGAGCGGTACCGCGCGCGCTC 481
               |||

```

RESULT 39
US-09-252-991A-6506
Sequence 6506, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6506
LENGTH: 477

```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-6506

Alignment Scores:
Pred. No.: 0.00128 Length: 477
Score: 118.50 Matches: 44
Percent Similarity: 37.23% Conservative: 26
Best Local Similarity: 23.40% Mismatches: 57
Query Match: 4.51% Indels: 61
DB: 4 Gaps: 8

US-10-088-045-2 (1-502) x US-09-252-991A-6506 (1-477)

Qy      305  AsnIleGlnThrThrLeuIleProAlaGlyGlyAlaAlaIleThrGluPheLysValAsp 324
               |||
Db      8  AGCATGAGACCAACCATCTTC-----GCCGTGC----- 37
               |||
Qy      325  ValProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAla 344
               |||
Db      38  -----CACGCGATCATGCGCGGCTCGGAGC----- 61
               |||
Qy      345  LeuGlyIleLeuLysValGluGlyGluAsnHisGluIleTyrSerHisLysGlnThr 364
               |||
Db      62  ---GACATCTCGCGGAGAGCGGCGTGAAGAACTGCGCGCTACCTGCGACCGAACATC 118
               |||
Qy      365  AspAlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLys 384
               |||
Db      119  GCCGCGCTGAGCTGCGGAGAGC----- 145
               |||
Qy      385  ThrProAlaProAlaAsnLeuGlnGlnIleLysAlaGlyLysAlaThrTyrAspSer 404
               |||
Db      146  -----AAGCGCAGCTGACGAGCGCGGACGAGATCTCTCGCTC 184
               |||
Qy      405  AsnGlyAlaAlaCysHisGlnProAspGlyLys-----GlyValProAsn 419
               |||
Db      185  AACCTGTGCGCTGCGGAGAGCGCGGAGGAGTACCGCTGTGCGCGCGGAGAC 244
               |||
Qy      420  AlaPheProPheLeuAlaAsnSerAspTyrLeuAsnAlaAspHisAlaArgAlaIleSer 439
               |||
Db      245  CTGACCAAGCC-----GGCGGTTCATCTAGCGTCCAGCTACGCCAGCTCCAAACA 298
               |||
Qy      440  IleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGluSerVal 459
               |||
Db      299  ACCATCGGCATGCGCGGAGGCGCAG----- 325
               |||
Qy      460  MetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThrTyrThr 477
               |||
Db      326  ATGCCCGCGGAGAGCATCTGCGGAGGAGGAGGAGTGCACATCTGCGCGCTACATC 385
               |||
Qy      478  LeuAsnSerPheGlyAsnLysGly 485
               |||
Db      386  TACACCTGTGCGCAACCAAGGCG 409
               |||

```

RESULT 40
US-09-252-991A-1456/c
Sequence 1456, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 1456
LENGTH: 1494
TYPE: DNA

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1456

Alignment Scores:
Pred. No.: 0.00762 Length: 1494
Score: 118.50 Matches: 77
Percent Similarity: 35.78% Conservative: 45
Best Local Similarity: 22.58% Mismatches: 138
Query Match: 4.51% Indels: 83
DB: 4 Gaps: 16

US-10-088-045-2 (1-502) x US-09-252-991A-1456 (1-1494)

```

QY 18 LeuMetLeuSerGlyCysSerAsnGlnAlaAspLysAlaAlaGlnProlys----- 34
DB 1385 TTACCGCTAGACAAAGTCTCG-----GCCGCTCGCCGGCTGGCGGTG 1341
QY 35 SerSerThrValAspAlaAlaAlaLysThrAlaAsnAlaAspAlaAlaSerGlnGlu 54
DB 1340 TCGGCTCGCGCGCGCGCGCGCGCGCTGCGCTG--GCCGCTCGCGAGGTGGCGAGAG 1282
QY 55 HisGlnGlyGluLeuProValIleAspAlaIleValThrHisAlaProGluValProPro 74
DB 1281 TACGACTACGAAGTCTG-----ATCGCCCGC 1258
QY 75 ProValAspArgAspHisProAlaLysValValLysMetGluThrValGluLysVal 94
DB 1257 CCGCTGGAC----- 1249
QY 95 MetAlaGluLeuAlaAspGlyValGluIleValGlnPheThrPheGlyGlyGlnValProGly 114
DB 1248 CTGGAGATGCTCGCGCTTTCAGCAGCTCGCGCTGCTGCTATGCGCGCAAGTCCCGGC 1189
QY 115 GlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSerAsnHisProAsp 134
DB 1188 GTGAGCTTGGCGCGCAAGCGAGGCTGGGTGGCTGCGGTGCTTACCAACAGCGTGGAC 1129
QY 135 SerLysMetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyAlaGlu 154
DB 1128 GAA-----CCGACCAACATCCACTGCAACCGCATCGGCTGCGATCGAGATGCAAGCG 1075
QY 155 AlaSerPheThrAla-----ProGlyHisThrSerThrPheSerPheLys 169
DB 1074 GTGCGCTACATTTCCGAGCCCGCGGTGACACCGCATCGGCTGCTTCACTTCAAG 1015
QY 170 AlaLeuGlnProGlyLeuThrValTyrHisCysAlaValAlaProValGlyMetHisIle 189
DB 1014 ACCCGAGACCGCGCGAGCTACTGCTGCTACACCGCATCTGATGAGC--AGCGAGCAGCTC 958
QY 190 AlaAsnGlyMetTyrGlyLeuIleLeuValGluProLysGluGlyLeu----- 205
DB 957 GCGCGCGGACTGTGCGCGCGCGCTGATCATCGAGAGCGTGAACCCACCGGTTCCGCCAC 898
QY 206 ProLysVal-----AspLysGluIleValMetGlnGlyAspPheTyrThrLysGly 223
DB 897 GAGAAAGTTCTCTGCTGAAGACTGCGACGTG----- 865
QY 224 LysTyrGlyGluGlnGly---LeuGlnProPheAspMet---GluLysAlaIleArgGlu 241
DB 864 -----GACGAGAGGGCGCGCTTACCCCTTCAAGGTGCGCGCGCGCGCTGA 811
QY 242 Asp-----AlaGluTyrValValPheAsnGlySer----- 251
DB 810 GGCACCGCGCGCGCGCTATTGACGATCAACGCAAGCATGTGCCGACATCGACCTGCCG 751
QY 252 ValGlyAlaLeuThrGlyLysAlaLeuLysAlaLysValGlyIleThrValArgLeu 271
DB 750 GCGCGCGAGATGCTCGCGGTGCGCTGTAAGCTGACACACCGCTACCTACCTGCTC 691
QY 272 PheValGlyAsnGlyLysProLeuThrSerSerPheHisValIleGlyGluIlePhe 291
DB 690 AACCTGCCCAAGCGCGAGGCGAGATC-----TATGCCATCGATGGCCATCCGGTG 640

```

```

QY 292 AspLysValHisPheGluGly-----GlyLysGlyGluAsnHisAsnIleGln 307
DB 639 GAACACCGCGCTTCCGAGGCGAGTACTGATCGGTCCCGGATCGCTGGAGCTGGCG 580
QY 308 ThrThrLeuIleProAlaGlyValAlaIleThrGlnPheLysValAspValProGly 327
DB 579 CTGAAGTCCCGAAGCGGGCAC--CGAGCTGTCTCGCGCGAGCGCGCTGAGCTGGC 521
QY 328 Asp 328
DB 520 GAC 518

```

Search completed: September 7, 2003, 08:08:11
Job time: 27346 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 7, 2003, 00:25:01 / Search time 375 Seconds
(without alignments)
3076.903 Million cell updates/sec

Title: US-10-088-045-2

Perfect score: 2626
Sequence: 1 MSKPTLIKTTIKALCALML.....NKGQSLADDVAKAKTKKEN 502

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1537136 segs, 114998732 residues
Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USFTO.spool/US10088045/rnat_04092003 083147_4078/app query.fasta_1.647
-DB=Published Applications_NA -QFMT=fastcap -SUFFIX=rmpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62
-TRANS-human40.cci -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10088045@cgn_1_1_244 @rnat_04092003 083147_4078
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELop=6 -DELext=7

Database : Published Applications NA:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/prodata/1/pubpna/US09C_NEW_PUB.seq:*
13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152.5	5.8	1125	14	US-10-127-032-31 Sequence 31, Appl

2	142	5.4	1533	10	US-09-738-626-1053	Sequence 1053, Ap
3	142	5.4	3309400	10	US-09-738-626-1	Sequence 1, Appl
4	140.5	5.4	1479	10	US-09-738-626-3280	Sequence 3280, Ap
5	139	5.3	1722	10	US-09-938-842A-1675	Sequence 1675, Ap
6	136.5	5.2	980	15	US-10-174-693-208	Sequence 208, Ap
7	132	5.0	1713	10	US-09-938-842A-1910	Sequence 1910, Ap
8	129	4.9	2230	12	US-10-361-460-78	Sequence 78, Appl
9	128.5	4.9	2297	14	US-10-115-563-17	Sequence 17, Appl
10	128.5	4.9	2297	14	US-10-115-563-27	Sequence 27, Appl
11	128.5	4.9	2297	14	US-10-115-563-28	Sequence 28, Appl
12	128.5	4.9	6295	10	US-09-880-107-2275	Sequence 2275, Ap
13	128.5	4.9	6295	14	US-10-115-563-13	Sequence 13, Appl
14	128.5	4.9	6295	14	US-10-115-563-26	Sequence 26, Appl
15	125	4.8	9025608	14	US-10-156-761-1	Sequence 1, Appl
16	119	4.5	1626	14	US-09-938-842A-1754	Sequence 1754, Ap
17	118.5	4.5	9025608	14	US-10-156-761-1	Sequence 1, Appl
18	117	4.5	2151	14	US-10-127-032-72	Sequence 72, Appl
19	117	4.5	8911	13	US-10-044-090-17	Sequence 17, Appl
20	114.5	4.4	1791	9	US-09-338-723A-5	Sequence 5, Appl
21	114.5	4.4	1791	13	US-10-080-233-1	Sequence 1, Appl
22	114.5	4.4	1791	13	US-10-080-233-1	Sequence 30, Appl
23	114.5	4.4	13842	10	US-09-861-289-30	Sequence 30, Appl
24	114.5	4.4	13842	10	US-09-860-846-30	Sequence 30, Appl
25	114.5	4.4	13842	11	US-09-988-384B-30	Sequence 30, Appl
26	114.5	4.4	13842	11	US-09-836-821-30	Sequence 30, Appl
27	114.5	4.4	36778	10	US-09-861-289-5	Sequence 5, Appl
28	114.5	4.4	36778	10	US-09-860-846-5	Sequence 5, Appl
29	114.5	4.4	36778	11	US-09-836-821-5	Sequence 5, Appl
30	114.5	4.4	37948	11	US-09-988-384B-5	Sequence 5, Appl
31	114.5	4.4	38506	11	US-09-793-708-19	Sequence 19, Appl
32	114.5	4.4	38506	12	US-10-201-365-1	Sequence 1, Appl
33	114.5	4.4	38506	12	US-10-160-539-19	Sequence 19, Appl
34	114.5	4.3	1938	14	US-10-156-761-4392	Sequence 4392, Ap
35	114	4.3	2670	13	US-10-002-600-44	Sequence 44, Appl
36	113.5	4.3	6026	10	US-09-995-749A-1	Sequence 1, Appl
37	113	4.3	7914	13	US-10-095-718-3	Sequence 3, Appl
38	113	4.3	27360	10	US-09-070-927A-164	Sequence 164, Appl
39	112.5	4.3	2464	12	US-09-960-706-639	Sequence 639, App
40	112.5	4.3	2464	14	US-10-205-823-290	Sequence 290, App
41	112.5	4.3	2481	9	US-09-853-386-89	Sequence 89, Appl
42	112.5	4.3	2481	12	US-10-101-510-139	Sequence 139, App
43	112.5	4.3	2481	14	US-10-211-239-8	Sequence 8, Appl
44	112.5	4.3	2481	14	US-10-288-222A-5	Sequence 5, Appl
45	112	4.3	1608	14	US-10-156-761-555	Sequence 555, App

ALIGNMENTS

RESULT 1
US-10-127-032-31
Sequence 31, Application US/10127032
Publication No. US20030113742A1
GENERAL INFORMATION:
APPLICANT: Whiteley, Marvin
APPLICANT: Banger, M. Gita
APPLICANT: Lory, Stephen
APPLICANT: Greenberg, Everett Peter
TITLE OF INVENTION: BIOFILM FORMATION
FILE REFERENCE: UIZ-070CP
CURRENT APPLICATION NUMBER: US/10/127,032
PRIORITY FILING DATE: 2002-04-19
PRIORITY FILING DATE: 2001-04-20
PRIORITY FILING DATE: 2001-04-20
PRIORITY FILING DATE: 2001-10-24
NUMBER OF SEQ ID NOS: 170
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 1125
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-10-127-032-31

OY		84	VAlValValLysMeGluThrValGluLysValMetAspLeuAlaAspGlyValGluTyr	103
Dd		175	GTCACACTTGCCCTGGAGGCTCAGACTGGGAGAAGCTCAAGATTTGCCGAATGCACAACG	234
OY		104	GlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGluGlyAsp	123
Dd		225	AAGACGTGGGGTTTCAATGGCATCTATTGGGGCGCAGCTTGtGGTGAAGAAAGTGTAT	294
OY		124	ThrlleGluValGlnPheSerAsnHisProAspSer	135
Dd		295	GACGTCACAGCTTGATGATGATAAACAATTTGGATGAATAGAACACTGTGCATGGCATGGC	354
OY		136	---LysMetProHiscAsnValAspPhe-----HisAlaIleThrGlyProGlyGlyGly	152
Dd		355	ATGAAGTTGGCGGCGCATGCTGTATGTGTGTGTCGCGCATCCAGATGGGCTGGGCGACAGC	414
OY		153	AlaGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGln	172
Dd		415	TGCTACCAGACTGTGACTGTGGCCAAATATGACAGCCACTTTGTGTACACCCGCACACT	474
OY		173	ProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGly	192
Dd		475	CATGGCCCTG-----ACAGTTTGCATGCGCTCACGTGCT	507
OY		193	MetTyrGlyLeuIleLeuValGluProLysGluGlyLeuProLysValAsp-----Lys	210
Dd		508	TTGGCGGGAGATATCATTTGTGGAA---GATGAAGCAACAGACAGCTGATCTGCACAGC	564
OY		211	GluTyrTyrVal-----MetGlnGlyAspPheTyrThrLysGly	223
Dd		555	GAGTTCGGTGTGGACGATATTCGGCTGTTTTTAATGATATCACCGCTTTTGAAGAAGCGT	624
OY		224	LysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGluAspAla	243
Dd		625	TCCCTTAGTAGAAGAACCTCCCC---CATCTTGGCTGTGTGGCGGATACCCCCACT	678
OY		244	GluTyrValValPheAsnGlySerValIleValIleThrGlyGluAsnAlaLeuLysAla	263
Dd		679	GCCAATGGCATTATCCCAATGCCCATCTTGATGACACACAGCCCGGGTTCCGTTCCGCGTG	738
OY		264	LysValGlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeu-----	280
Dd		739	CTCAACGGCTCCAAATATGCGGTTCTAT-----AACTTGGCGTTTCA	780
OY		281	--ThrSerSerPheHisValIleGly-----GluIlePheAsp	292
Dd		781	GACAGCGGCACCTTSCAATGATATTCSSAGCATTCGCGTTGCTGGATGAACCTCAAGAC	840
OY		293	LysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThrThrLeuIlePro	312
Dd		841	CGCACACACTTGGCTATTGGCCCAAGCGCAGCGGTGGAAATGTCGTGACATGCAGCCC	900
OY		313	AlaGlyGlyAlaAlaIleThr-----GluPheLysValAspValPro	326
Dd		901	-----GGCGAGAGACTACCTTGGGAATCTGTAGTTTGAAGACAATACAGCGGCTCCT	954
OY		327	GlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGly	346
Dd		955	GATGATGATGTCGTGTCGCCGATTTCCGCAATGCAGATCTCTTCGAG-----	999
OY		347	IleLeuLysValGluGlyGluGluAsnHisGluIleTyrSerHisLysGlnThrAspAla	366
Dd		1000	CTGCTCAACATCAGCCGCCCTTCCGAT-----GATGCT	1032
OY		367	ValTyrLeuPro-----GluGlyAlaProGlnAlaIleAsp	378
Dd		1033	GGCGAAGACACTTCTTGGCGGAGCTGCTGGTGAATATCACCGAACTGACGTCACTCAT	1092
OY		379	ThrGlnGlu-AlaProLys-----ThrProAlaProAlaAsnLeuGlnGluGlnIleLys	396
Dd		1093	GCGATGAAGCAGACTTCATGATGAACAACCTTCTCACAAGAGATC-----	1138
OY		396	sAlaGlyLysAlaThrTyrAspSerAsnCysAlaAla	408

[illegible]

```

Oy      175 LeuTYValTYRHISCyAAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyr 194
Db      3169855 ACCTHACTTTACCAT-----TCACAAACCGAGCTGCAGCTTGATCGGGCCCTCAC 3169905
Oy      195 GlyIleLeuValGluProLysGluGlyLeuProLysValAspLysGlyTyrVal 214
Db      3169906 GCCCACTGATCATCCTGTGACCAGCCGGAAGAAGCTGAGAACGAGACGTGATGGAGCATTC 3169965
Oy      215 MetGlnGlyAspRhePheTyrThrLysGlyLysTyrGlyGlnGlnGlyLeuGlnProPheAsp 234
Db      3169966 GTGCTCGACGACCTGGGTGCAT-----GGCATTCAGGGCACTCCGACGATGAG 3170013
Oy      235 MetGluLys----- 237
Db      3170014 CTCGACAAGCTCACCGGAAATGGGTTCGGGSSACCATTAACGGAGAGATGGGAATTGGAGCT 3170073
Oy      238 -----AlaIleArgGluAspAlaGluTyrValValPheAsnGlySerValGlyAlaLeu 255
Db      3170074 CACGCCAGATGATGACGAGCGACCCCGGACCGGGGACTGGGGCGGGGATGTGCGGAGATGTG 3170133
Oy      256 -----ThrGlyIleAsnAlaLeuValAla 263
Db      3170134 ATGTAATCCGCACTACCTCATCAACGAGAGTATCCCGCGTGTCAACCGACCTTGAGAGGCT 3170193
Oy      264 LysValGlyGlyThrValArgLeu---PheValGlyAsnGlyGlyProAsnLeuThrSer 282
Db      3170194 CGCCGGGCGCACAGGCCCGCGCTGGGTGTTATCAACTCCGGCGGTGACACCAT- 3170247
Oy      283 SerPheHisIleIle-----GlyGlyLeuPheAspLysValHisPheGlnGlyLys 300
Db      3170248 ---TTCAGAGTGGCGCTCGGTGGTGTACCGGATGACTGACACCCACCGAGCGCTTCCT 3170304
Oy      301 GlyLeuAsnHisValIleGlnThrThrLeuIleProAlaGlyAlaAlaIleThrGlu 320
Db      3170305 GTCGACCGCTGGGAGACCGGACATGATCTACCTGATGAGGCGAG----- 3170349
Oy      321 PheLysValAspValPro-----GlySerTyrValLeuValAspHisIlePhe 337
Db      3170350 ---CGTGTGACGTGAGATCATCTCGGCGACCGGACATTCCTCGACGGCTTTGGCG 3170406
Oy      338 ArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGlnGlyIleAsnHisGlu 357
Db      3170407 GTGGGTAAAGACGACCGGCGCTTGCGCGTCAATCCGACCGCGCGGCGGACGGCCCCCG 3170466
Oy      358 -----IleTyrSerHisLysGln 363
Db      3170467 CCCGATGTCGACTTCCCAGATGTGTCGTCACACCGGACGTGCTGTCTCTCCTDAAGCA 3170526
Oy      364 ThrAspAlaValTyrLeuProGlnGlyAlaPro 374
Db      3170527 GCAGACCGTGCACCTGTGCGCGAGGACACCA 3170559

RESULT 4
US-09-738-626-3280
; Sequence 3280, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUKIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIO
; APPLICANT: KEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484

```

	PRIOR FILING DATE: 1999-12-16	
	PRIOR APPLICATION NUMBER: JP 00/159162	
	PRIOR FILING DATE: 2000-04-07	
	PRIOR APPLICATION NUMBER: JP 00/280988	
	PRIOR FILING DATE: 2000-08-03	
	NUMBER OF SEQ ID NOS: 7059	
	SOFTWARE: PatentIn ver. 3.0	
	SEQ ID NO 3280	
	LENGTH: 1479	
	TYPE: DNA	
	ORGANISM: Corynebacterium glutamicum	
	US-09-738-626-3280	
	Alignment Scores:	
	Pred. No.: 1,65e-05 Length: 1479	
	Score: 140.50 Matches: 77	
	Percent Similarity: 34.53% Conservative: 58	
	Best Local Similarity: 19.69% Mismatches: 180	
	Query Match: 5.35% Indels: 78	
	DB: 10 Gaps: 13	
US-10-088-045-2 (1-502) x US-09-738-626-3280 (1-1479)		
Oy 44 ThrAlAsnAlAaspAsnAlAlaserGlnGlunHISgInglYglLueProValIleAsp 63		
Db 24 ACGATTTCGTCTGGCGGGCTGTACTTGCAGCCGCCGGGGCCGTG-CGGCTGCACCA 82		
Oy 64 AlAlIleValThriSAlaProgluValProPropoValaAspArgPheHisProAlaLys 83	::::	::::
Db 83 GCGACCCTGAAACCGCTGCTCG-GCACAGAAGTCTCCTTGCCCACTCCACCCC 144	::::	::::
Oy 84 ValValalLymecLutThrValglulLysValMetArgLue----- 97	::::	::::
Db 142 ACTGCGCTGGTAGCGCGAGCGGTGGCGGACATGAACGCCGCCGCCCTCTCCCTGAT 201	::::	::::
Oy 98 AlaaspGLValglutyrqLnPhetrPrphrPheglYglValAlProglYlmetIle 117		
Db 202 ATGGCGGCAATGAAGCCAAGACGTGGGATACGTCTTGACACCGGGATCGGCATT 261		
Oy 118 ArgValArgLugluYasPrThrllegluValGlnPheSerAnhis---ProaspSerLys 136		
Db 262 GAGGCACCGCGCGAGCGCTCTCCAGGTCGATATACAAATGAACATCGCGAAGAAC 322		
Oy 137 MetPro----HisenValasPhehisAlaAlathrgLYProglYglYglValGlu 154	::::	::::
Db 322 TCCATCCATGGCATGCGCATCGCACTCCACAAGCGCGAGCGGTGGCCGCGATGACC 381	::::	::::
Oy 155 AlaserPheThralProglYhiserThrhPheSerPheLysAlaleuGlnProglY 174		
Db 382 CAGGACCCCAATTGAACCTGGCGAGCTTTCTTCCTATGTTTGAAGTCCCCACGGTGGC 441		
Oy 175 LeuTYValTYTrhisCYsaAlavalAlaProvalGlymeChisilalaasnGLymetYr 194		
Db 442 ACCTACTTCTACCAT-----TCCACAACCGGCGCTGAGCTGATTCGGGCTCCAC 492		
Oy 195 GlyLeuIlleLeuValgluProLysegluYlLeuProLyValaAspLysgluTYrTYVal 214	::::	::::
Db 493 GCCCCATGATCATCCGTGACCCGGAAGACCTGAGACAGCAAGCTCGAGTGGACATC 552	::::	::::
Oy 215 MetGlnGLyasPheTYrThriYselYlYsTYrGLYglGlnGLYleuGlnProPhaAsp 234	::::	::::
Db 553 GTGCTCGAAGCATGGTGCAT-----GGCATTCAGGCACTCCGACGATGAG 600	::::	::::
Oy 235 MetGluLys----- 237	::::	::::
Db 601 CTCGACAACCTACCGGAATGGGTTCGGGCGACCAATAACGGAAGATGGCAATGGAGCT 666		
Oy 238 -----AlaileArGLuAspAlagluTYrValValPheasnGLYserValglYAlaleu 255		
Db 661 CACGGCAATGATGACGGACCCCGGACCGGGATCTGGGCGGGATGTCGGCATGTG 720		
Oy 256 -----ThrgLYgluAsnAlaleuLYsAla 263	::::	::::

Db 721 ATGATCCGACATCTCATACAGGAGTATCCCGCTGCTACCGGACTTGGAGGCT 780
 Qy 264 LysValAlaGluThrValArgLeu---PheValGlyAsnGlyProAsnLeuThrSer 282
 Db 781 CGCGGGGCGCAAGAGCGCGCTCGGGTTTATCAATCCGGCGGTGACACCATC----- 834
 Qy 283 SerPheHisValIle-----GlyGluIlePheAsnLysValHisPheGluGlyLys 300
 Db 835 ---TTCAAGGTGGCGCTCGGTGTACCGGACGTGACCTGCACACCGAGCGGCTTCCCT 891
 Qy 301 GlyGluAsnHisAsnIleGlnThrLeuThrLeuIleProAlaGlyAlaAlaIleThrGlu 320
 Db 892 GTCCAGCCCTGGGAGACCGAATGATCTACTGTGATGGCGAG----- 936
 Qy 321 PheLysValAspValPro-----GlyAspTyrValLeuValAspHisAlaIlePhe 337
 Db 937 ---CGTTCGACGTCGAGGTGATCTCTGGGAGCGGACATCTTCCGTCACGGCTTTGGCG 993
 Qy 338 ArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsnHisGlu 357
 Db 994 GTGGGTAAAGACGACCGCGCTTTCGCGGTATCCGACCGCGCGGCGGAGCGCGCCGC 1053
 Qy 358 -----IleTyrSerHisLysGln 363
 Db 1054 CCCGATTCGACTTCCCGAGTTGTCTCCACCGGACTGCTTCTGCTCCGTAAGCCA 1113
 Qy 364 ThrAspAlaValTyrLeuProGluGlyAlaPro 374
 Db 1114 GCAGACCGTGCATCTCTGCGGAGGCGCACCA 1146

RESULT 5

US-09-938-842A-1675
 ; Sequence 1675, Application US/09938842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Krepes, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OR INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRIP1300-3
 ; CURRENT APPLICATION NUMBER: US/09/938, 842A
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227, 866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264, 647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300, 111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 1675
 ; LENGTH: 1722
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-938-842A-1675

Alignment Scores:

Pred. No.: 3e-05 Length: 1722
 Score: 139.00 Matches: 101
 Percent Similarity: 33.48% Conservative: 54
 Best Local Similarity: 21.81% Mismatches: 137
 Query Match: 5.29% Indels: 171
 DB: 10 Gaps: 26

US-10-088-045-2 (1-502) x US-09-938-842A-1675 (1-1722)

Qy 80 HisProAlaLysValValValMetGluThrValGluLysValMetArgLeuAlaAsp 99
 Db 49 CACTCGCGCGTCCGGCGGAGTAGATCAACGCGGAG----- 87
 Qy 100 GlyValGluTyrGlnPheTrp-----ThrPheGly 109

Db 88 ---GTTGAGTAAAGTACTGCTGCGCGGATTTGAAGAGAAATGTTATGGCCATCAAC 144
 Qy 110 GlyValAlaProGluGlnMetIleArgValArgGluLysAspThrIleGluValGlnPhe 129
 Db 145 GCGCCAGTTTCCAGGCGCCAAAGATGACAGCATGCGCGGAGACACGTCATATTC----- 198
 Qy 130 SerAsnHisProAspSerLysMetProHisAsn-----ValAspPheHisAla--- 145
 Db 199 ---CACGTCTCAACAAATCTCCACGGAAGGTGTGTATTCATTGGACAGCGCAT 252
 Qy 146 -----AlaThrGlyProGluGlyValAlaGluAlaSerPheThr 158
 Db 253 CGTCGAAAGGACCTCATGAGTGTAGAGCAGCATGAGTGCACCGCTCTATTAAAT 312
 Qy 159 AlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrValTyr 178
 Db 313 ---CTGGCGAGACTTTCATCTTACAACTTATGTCGATGAAGCGGGAACAATTTTAC 369
 Qy 179 HisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeu 198
 Db 370 GAT-----GACACATACGGAATGCACAAAGATCATCGGAGCTATACGGAATGTATA 420
 Qy 199 Val---GluProLysGluGlyLeuProLysValAspLysGluTyrTyrValMetGlnGly 217
 Db 421 GTAAGATCACCAAAAGAGAGATTAA---ATTAGATGAGAAAGTTCAATCTTGGTCACT 477
 Qy 218 AspPheTyrThrLysGlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLys 237
 Db 478 GACTGTGGGACCAAAAGCATTCACGGCCAA----- 507
 Qy 238 AlaIleArgLysAspAlaGluTyrValValPheAsnGlySerValGlyAlaLeuThrGly 257
 Db 507 ----- 507
 Qy 258 GluAsnAlaLeuLysAlaLys-----ValGlyGluThrValArgLeuPheVal 273
 Db 508 GAACGTCTCTTCTTCT 567
 Qy 274 ---GlyAsnGlyLysProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAsp 292
 Db 568 AATGGAAGACGACATTCATTTGTTCTTCAAGACCGTAT-----TTTAAAC 612
 Qy 293 LysValHisPheGluGlyLys-----GlyGluAsnHisAsnIle 306
 Db 613 AAA-----GGAGGAGAGAAAGATGTATGCACGTTTAAAGAAATGATCACTGT 660
 Qy 307 ---GlnThrThrLeuIlePro-----AlaGlyGly 315
 Db 661 GCACCTCAAACTTTCGAGTCGAACCAATAGATGATACCGTTCGATCGCTAGCACA 720
 Qy 316 AlaAlaIleThrGluPheLysValAspValProGlyAspTyrValLeuValAspHisAla 335
 Db 721 ACTGCTCTGCTTCTCTCTCAACTTGGCGGTTCACAGACACCGAGTGGT----- 771
 Qy 336 IlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuLysValGluGlyGluAsn 355
 Db 772 -----GTTGAGCTGAGCGGAC 789
 Qy 356 His-----GluIleTyrSerHisLysGlnThrAspAla 366
 Db 790 TACGTGCAACGTTTCCACCGTCACAGACATTCAGCTTATTC---GGCGAACTTATTC 846
 Qy 367 ValTyrLeuProGluGlyAla----- 373
 Db 847 GTTCTCTTAAACCAACGCACTTCATCAAGAAAGTACTGATCTCCGTGGCGTTTCT 906
 Qy 374 -----ProGlnAlaIleAsp-----ThrGln 380
 Db 907 GCGCGAAGAACCAAAATCTCTCAAGACATCACCGTGAATAATTACGTGATGCCACGTAG 966
 Qy 381 GluAlaProLysThrProAlaProAlaAsn---LeuGlnGluGlnIleLysAlaGlyLys 399
 Db 967 TCACGCCCATCTCATCACCACCGGTGACTTCATCTGGAACGACACAGATCGAGCAAA 1026

```

Oy 400 AlathrTyraSerAsnCyAlaAlaCysHISGInProAspGlyLysGlyValProksn 419
Db 1027 AGCTTCGAGAGAAAGATTTGGCCGCT-----AAAGGATATCCAAA 1068
Oy 420 AlaHeProPLeuAlaAsnSerAspTyLeu-AsnAlaAspHisAlaArgAlaLase 439
Db 1069 -----CGCCGGAGAAATCAATGACACGCTAAATCTCTCAACACACAGA----- 1114
Oy 439 rIleValAlaAsnGlyLeuSerGlyLysIleThyValAsnGlyAsnGlnTyrgLysSerVa 459
Db 1115 -----ATCTCTACGAGATTACAGAAATGTCATACACAGCTTCA 1158
Oy 459 lmePro 461
Db 1159 TTATCCG 1165

RESULT 6
US-10-174-693-208
; Sequence 208, Application US/10174693
; Publication No. US20030131373A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000, 10035
; CURRENT APPLICATION NUMBER: US/10/174,693
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 208
; LENGTH: 980
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-174-693-208

Alignment Scores:
Pred. NO.: 2,52e-05 Length: 980
Score: 136.50 Matches: 71
Percent Similarity: 34.51% Conservative: 46
Best Local Similarity: 20.94% Mismatches: 129
Query Match: 5,204 Indels: 93
DB: 15 Gaps: 14

US-10-088-045-2 (1-502) x US-10-174-693-208 (1-980)
Oy 78 ArgAspHisProAlaValValValValLysMetGluThrValGluLysValMetArgHeu 97
Db 27 CGTAACCTACACATTCATGAGTAAAGAAATACCAG-----AGACTG 71
Oy 98 AlaAspGlyValGluTyrgInPheThrPheGlyGlyGlnValProGlyGlnMetIle 117
Db 72 TGTTCG-----AGCAAGCCCATCTGACCGTGAAGGAGATTTCCCGGACCCCATCTC 125
Oy 118 ArgValArgGlyGlyAspThrIleGluValGlnPheSerAsnHisProAspSerLysMet 137
Db 126 TATGCTAGGAGAAATGACACCGCTGCTCGTAGGGCTCTTAAC-----CGTCTC 173
Oy 138 ProHisAsnValAspPheHis-----AlaAlaThr 147
Db 174 AAATCAAAATGTCACCATTCGATGGATATCCGGCAGTTGAGAGACGGGGTGGCCGAC 233
Oy 148 GlyProGlyGlyGlyAlaGluLaserPheThrAlaProGlyLysIleThrSerThrPheSer 167

```

```

Db 234 GGGCCACGATACATTTACCAATGCCGATC-----CAGCCGGGCGAAAGCTATGTGTACAT 290
Oy 168 PheLysAlaLeu---GlnProGlyLeuTyrValTyriScyAlaValAlaProValGly 186
Db 291 TTACACATCAAGGGGCAAGGGGACCCCTCTGTGATGATCAACATTA----- 338
Oy 187 MetHisIleAlaAsnGlyMetTyrgLysLeuIleLeuValGluProLysGlyGlyLeu--- 205
Db 339 CTGTGCTCAGGGGCAACCTGACAGGACCATTTGATCTTGGCCAAAGCGGTGTCCA 398
Oy 206 -----ProLysValAspLysGlyTyrgTyrgValMetGlnGlyAspPheTyrgLys 222
Db 399 TACCTTTCCCTTAACCCCAAGAGATTTGTCTATATGGCGAATGGTGAATCT 458
Oy 223 GlyLysTyrgLys-----GluGlnGlyLeuGlnProPheAspMetGlu 236
Db 459 GATACAGAGGTGTGATCACTCAAGCATCAAGTCCGATTAAGCAGC----- 506
Oy 237 LysAlaIleArgIleAspAlaGluTyrgValValPheAsnGlySerValGlyAlaLeuThr 256
Db 507 -----AATGCTCCGATGCTACACAGATCAATGCGCATCCAGGGCCAAATGCC 554
Oy 257 -----GlyGluAsnAlaLeuLysAlaLysValGlyGluThrValArgLeu 271
Db 555 AATGCCCTTCCAGGGGTGATTTACCTGCTGTGTAGAGTGCAGAGAAATACATCTCG 614
Oy 272 PheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePhe 291
Db 615 CGAATCATCAACGCTGGCTCAATGAGAGCTTTCTTCAAGATGCCGGGAC----- 668
Oy 292 AspLysValHisPheGluGlyGlyLysGlyGluAsnHisAsnIleGlnThrLeuIle 311
Db 669 -----CAGTCGACCATCGTG 683
Oy 312 ProAlaGlyAlaAlaIleThrGluPheLysValAspValProGlyAspTyrgValLeu 331
Db 684 GAGCTCAGCGCACTTCAAGCTTTTCAAGCCACAGC----- 725
Oy 332 ValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeuGlyIleLeuValGlu 351
Db 726 -----ATCGCATTCACCT 740
Oy 352 GlyGluGluAsnHisGluIleTyrgSerHisLysGlnThrAspAlaValTyrgLeuProGlu 371
Db 741 GGCAGAACCAACATGCGCTCATCTCCACGACAGACAGCTGTGCAATGATCATGTGCC 800
Oy 372 Gly-----AlaProGlnAlaIleAspThrGlnGluAlaProLysThr 385
Db 801 GCCTCCCTTTTATGACTCCCGATGCGCTCGACAAATGACCGGACCGCACCA 857

RESULT 7
US-09-938-842A-1910
; Sequence 1910, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepes, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1910
; LENGTH: 1713

```


Qy	161	GLYHISTSRSTHrPhSeSerPheValAlaLeu---GlnProGlyLeuTyrValTyrHis	179
Db	506	GGGCAAGATTACGCTGACATCAACCCCTCGGGGACCGGGACCGCTGGTGGAC	565
Qy	180	CysAlaValAlaProValIGlyMetHisLeuAlaAsnGlyMetTyrGlyLeuIleVal	199
Db	566	GGGCACATCTCC-----TGGCTGGCGCGCACCGCTCTACGGGCCCTCGTCAATC	613
Qy	200	GluProLysGlyGlyLeu-----ProLysValAspLysGlyTyrTyrValMet	215
Db	614	CTGCCCAAGCTCCGGGCTCCCTACCCGTTCCCGGCGCCCTACAGAGAGTCTCCCTCAATC	673
Qy	216	GlnGlyAspPheTyrThrLysGlyLysTyrGlyGlnGlyLeuGlnProPheAspMet	235
Db	674	TTCCGTGATGATGGTGGCTGGC-----GACACG	700
Qy	236	GluLysAlaIleArgGlu-----AspAlaGlu	244
Db	701	GAGGTGGTGAATCAAGACAGCGCTTACGTCGGCGCTGGCCCAATCTCTGACGCCAC	760
Qy	245	TyrValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAla-----	260
Db	761	ACCATC-----AACGGCTCGCATAGCGGCTCTCAACTGCTCTGCCAAGACACGTAC	814
Qy	261	---LeuLysAlaLysValGlyGlnThrValArgLeuPheValGlyAsnGlyGlyProAsn	279
Db	815	AAGCTGAAGGTGAAGCCCGGAGAGACGTACATGCTCGGCTCTATCAACGGCGCTCAAC	874
Qy	280	LeuThrSerSerPheHisValIleGlyGlnIlePheAspLysValHisPheGlyGly	299
Db	875	GACGAGCTCTTCTTCCGCTGGCCACCACTCGCTCAAGCTGTGAGAGTCGACGCCCTC	934
Qy	300	LysGlyGluAsnHisAsnIleGlnThrThrLeuIleProAlaGlyGlyAla-----Ala	317
Db	935	TACGTCAAGCCCTTCAACGCTGCACACGCTGCTCATCGCGCGGCGGACACCAACATG	994
Qy	318	IleThrGlyLysPheLysValAspValProGlyAspTyrValLeuValAspHisAlaIlePhe	337
Db	995	CTGCTCGCGCGGCAACCGCTCTACCCGGGCGGCACTACATCTCGCGCGGCTCAAC	105
Qy	338	ArgAlaPheAsnLysGlyAlaLeuGlyLysLeuLysValGlnGlyGlnGluAsnHisGlu	357
Db	1055	TTCACCGCGGACGCGGCGGCACTTGCACAAACACACCGCTCGCGGCACTCTGCAGTACAG	1111
Qy	358	IleTyrSerHisLysGlnThrAspAlaValTyrLeuProGlyGlyAlaProGlnAlaIle	377
Db	1115	CTGTAC-----	1122
Qy	378	AspThrGlnGlnAlaProLysThrProAlaProAlaAsnLeuGlnGlnIleLysAla	397
Db	1121	-----CCCGACCGCGCGCGGCTCCGC-CTCCGGGGAGAGCTTCACGA	1166
Qy	398	GlyLysAlaIleThrTyrAspSerAsnCysAlaIaAsyHis-----	410
Db	1165	GGCCCTGCGCTTACAGACGACCCCTGCGGAGCTCAACGACACCAACTTCTGGGCA	1222
Qy	411	-----GlnProAspGlyLysGlyAlaProAsn-AlaPheProPhe	424
Db	1225	CTTCACGCGCAAGCTCCGACGCTCGCGACCGCGGGATACCGGGGCGCTGCGCGGAC	1288
Qy	424	AlaIaAsnSerAspTyrLysAsnAla-----AspHisAlaIaArgAlaIaSerI	440
Db	1285	GGTGACACGGCGGTTCTTTCGCGGTGGGCTCGGACGACCGCTGCCCGGCAACGC	1344
Qy	440	ValAlaAsnGlyLysSerGly-----LysIleThrValAsnGlyLysGlnTyrGly	458
Db	1345	CACGTGCCAGGGCCCCACCAACACACAGAGTTCCGGCGGTCTGTACACACAGTCTCTT	1404
Qy	458	ValMetProAlaIleAlaIle	465
Db	1405	CGTGTCTCCCAACAGCGGCTG	1426

```

US-10-115-563-17
; Sequence 17, Application US/10115563
; Publication No. US2003008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; Greengard, Judith S
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel
; STREET: 10666 No. US2003008307Alth Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/115,563
; FILING DATE: 02-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,488
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Flitig, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: 449.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_difference
; LOCATION: replace(1614,"")
; OTHER INFORMATION: /label= N
; /note= "wherein "N" is a guanine in a Factor V
; normal allele and an adenine in a Factor V mutant
; allele"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1614_
; OTHER INFORMATION: /note= "Nucleotide position 1614
; below corresponds to nucleotide position 1691 of
; SEQ ID NO 13."
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-115-563-17
Alignment Scores:
Pred. No.: 0.00647 Length: 2297
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55
Best Local Similarity: 20.05% Mismatches: 145
Query Match: 4.89% Indels: 139
DB: 14 Gaps: 21
US-10-088-045-2 (1-502) x US-10-115-563-17 (1-2297)
Cy 63 AsphalalevalThrhAslaprogluValProbrotoValAspaq----- 78
:::|||||:::|||||:::|||||:::|||||:::
1085 GAAGTCATTGGGACACTATGCCCTGTATACCACCGCATATATGGCATAAAATATACAGTCT 1144

```



```

QY 79 -----AspHisProAlaIysValValIlys 87
Db 1145 CAGCATTGGATATTTCTCAACCAATTGAAAAACATTATAGAAAGTTATGACACA 1204
QY 88 Met-----GluThrValGluIysValMetArgLeuAlaAsp 99
Db 1205 CAGTACGAGATGATGCTTCACCAACATACAGTATCCCAATATGAAA---GAAGAT 1261
QY 100 GlyValGluIysGlnPheTrpPheGlyGlyGlnValProGlyGlnMetIleArgVal 119
Db 1262 GGG-----ATTGGGTCCTATATTCAGAGCC 1288
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerIys 136
Db 1289 CAGGTGAGAGACACACTCAAAATCGTTCAAAAATATGGCCAGCCGCCCTATAGACTT 1348
QY 137 MetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyGluAlaIleSer 156
Db 1349 TACCTTCATGAGTGAACCTTCGCTTATGAAGATGA-----GTCAACTCTTCT 1399
QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheIysAlaLeuGlnProGly---Leu 175
Db 1400 TTACCTCA---GGCAGGAACAACACCATG---ATCAGACGAGTTCAACAGGGGAAACC 1453
QY 176 TyrValTyr-----HisCys 180
Db 1454 TATACTTATATGAGAAACATCTTAGAGTTGATGAACCCACAGAAATGATGCCAGTGC 1513
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
Db 1514 TTAAACAGACACATCTAGCTGACGTGACATGATGACATGACATGCGCTCTGCTATATA 1573
QY 195 GlyLeuIleLeuValGluProIysGluGlyLeuProIys-----Val 208
Db 1574 GAGCTACTTATCTGTAAGACAGATCCCTGACAGCGGACGAATACAGAGGACGA 1633
QY 209 AspIysGluTyrTyrValMetGlnGlyAspPhe---TyrThrIysGlyIysTyrGlyGlu 227
Db 1634 GACATGGAACAGACAGGCTGTGTTGGCTGTTGATGAGAACAAAAGCTGTACCTTGAG 1693
QY 228 GlnGlyLeuGlnProPhe-----AspMetGlnIysAlaIleArgGluAspAlaGluTyr 245
Db 1694 GACACATCAACAAGATTGTGTGAATACTCGATGAGTGAAAGCTGATGACCCCAAGTTT 1753
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyGluAsnAlaLeuIysAlaVal 255
Db 1754 -----TATGATCAACATC 1768
QY 266 GlyIuThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerPheHis 285
Db 1769 ATGAGCATATCAATGCTATGTC-----CCTGAGACATTAAC 1807
QY 286 ValIleGlyGluIlePheAsp----- 292
Db 1808 ACTCTTGATTCGCTTGTATGACACTGTCAGTGCAGCTTGTAGTGTGGGACCCAG 1867
QY 293 -----LysValHisPheGluGlyGlyLys-----GlyGluAsnHisAsn 305
Db 1868 AATGAATTTTGCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 1927
QY 306 IleGlnThrThrLeuIleProAlaGlyAlaAlaIleThrGluPheIysValAspVal 325
Db 1928 GACACTTGACCTCTTCCCATGCGTGAAGATCTGTGACG----- 1969
QY 326 ProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnIysGlyAlaLeu 345
Db 1970 -----GTCAATGATGAT-----AATGTTGAACCTGG 1996
QY 346 GlyIleLeuIysValGluGlyGluAsnHisGluIleTyrSerHisIysGlnThrAsp 365
Db 1997 ATGTTAACCTTCATGAATCTAGTCCAAAGAAAGCTGAAGTTCAGGGAT 2056

```

```

QY 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProIysThr 395
Db 2057 GTTAAATGATATCCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2116
QY 386 ProAlaProAla-----AsnLeuGlnGluGlnIleIysAlaGlyLys 399
Db 2117 ACAGTATGCTTACACGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 2176
QY 400 AlaThrTyrAsp 403
Db 2177 GCTGACTATGAT 2188

RESULT 10
US-10-115-563-27
; Sequence 27, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; GREENGARD, Judith S
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel
; STREET: 10666 No. US20030008307A1th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10115,563
; FILING DATE: 02-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,488
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: 449.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1614
; OTHER INFORMATION: /note="Nucleotide position 1614
; below corresponds to nucleotide position 1691 of
; SEQ ID NO 13"
; SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-10-115-563-27

Alignment Scores:
Pred. No.: 0.00647 Length: 2297
Score: 128.50 Matches: 85
Percent Similarity: 33.02% Conservative: 55
Best Local Similarity: 20.05% Mismatches: 145
Query Match: 4.89% Indels: 139
DB: 14 Gaps: 21

```


Pred. No.: 0.000647 Length: 2297
 Score: 128.50 Matches: 85
 Percent Similarity: 33.02% Conservative: 55
 Best Local Similarity: 20.05% Mismatches: 145
 Query Match: 4.89% Indels: 139
 DB: 14 Gaps: 21

US-10-088-045-2 (1-502) x US-10-115-563-28 (1-2297)

Qy Aspa1a1leValThrh1sa1aPProgluValPProProValaAspArg----- 78
 Db 1085 GAAGTCATTGGAGCTATGCACTGTATATACCAAGATATGCAAAATAATACAGTCT 1144
 Qy 79 -----Asph1sProalalyValaValaValys 87
 Db 1145 CAGCATTGGATATTTCTCAACCAATGGAAACATTATAGAAAGTTATATACACA 1204
 Qy 88 Met-----GluThyValGluValMetArgLeuAlaAsp 99
 Db 1205 CAGTACGAGATGATGCTTCACCAACATACAGTATCCCAATATGAAA---GAAGAT 1261
 Qy 100 GlyValGluValGlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgVal 119
 Db 1262 GGG-----ATTGGGTCTCTATATACAGACC 1288
 Qy 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136
 Db 1289 CAGGTACAGACACACATCAAAATGCTTCAAAAATATGGCCAGCCGCCCTATAGCAT 1348
 Qy 137 MetProHisAsnValaPhePheHisAlaIleThrGlyProGlyGlyAlaGluAlaSer 156
 Db 1349 TACCCTCATGAGAGTCTTCCTGCTTATAGAAATGAA-----GTCAACTCTTCT 1399
 Qy 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu 175
 Db 1400 TTAACTCTA---GCCAGAAACAACACATG---ATCAGAGAGTTCAACAGGGGAAACC 1453
 Qy 176 TyrValTyr-----HisCys 180
 Db 1454 TATACTATAAGTGAACATCTTAGAGTTGATGACCCACAGAAATATGATGCCCATGTC 1513
 Qy 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
 Db 1514 TTAAACAGACCATCTACAGTACGTGACATCATGAGACATCGCTCGGCTATA 1573
 Qy 195 GlyLeuIleLeuValGluProLysGlyGlyLeuProLys-----Val 208
 Db 1574 GGACTACTTCTTAATCTGTAAGACAGATCCCTGGACAGGCAAGATACAGGGGACGA 1633
 Qy 209 AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyGlu 227
 Db 1634 GACATGCAACAGCAGCGGTGTTGCTGTGTTGATGAGAACAAACCTGATCCTGAG 1693
 Qy 228 GlnGlyLeuGlnProPhe-----AspMetGluLysAlaIleArgGluAspAlaGluTyr 245
 Db 1694 GACAACATCAACAAATTTTGAAATATCTGATGAGTGAACGTGATGACCCCAAGTTT 1753
 Qy 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyLysAlaLeuValaValaVal 265
 Db 1754 -----TATGATATCAACATC 1768
 Qy 266 GlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerPheHis 285
 Db 1769 ATAGACATATCAATGCTATGTC-----CCTGAGACATAACT 1807
 Qy 286 ValIleGlyGluIlePheAsp----- 292
 Db 1808 ACTCTTGATTTGCTTGTGATGACATGTCAGTGCAGTCTCTGATGAGGACCCAG 1867
 Qy 293 -----LysValaHisPheGlnGlyGlyLys-----GlyGluAsnHisAsn 305
 Db 1868 AATGAATTTTGAACATTCATCTGAGCACTCATTCATCTATGAAAGAGCATGAG 1927

Qy 306 IleGlnThrThrLeuIleProAlaGlyAlaAlaIleThrGluPheLysValaAspVal 325
 Db 1928 GACACCTTGACCTCTCCCATGCGCGAGAACTGTGTACG----- 1969
 Qy 326 ProGlyAspTyrValLeuValaPheHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345
 Db 1970 -----GTCAACATGAT-----AATGTGAACTTGG 1996
 Qy 346 GlyIleLeuLysValaGluGlyGluAsnIleGlyLysSerHisLysGlnThrAsp 365
 Db 1997 ATGTTAATCTTCATGATATTTAGTCCAAAGAACAAAGCTGAGCTGAATTCAGGAT 2056
 Qy 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThr 385
 Db 2057 GTTAAATGTATCCCATGATGATGAAAGACTCATATGATTTTGAACCTCCAGAAATCT 2116
 Qy 386 ProAlaProAla-----AsnLeuGlnGluGlnIleLysAlaGlyLys 399
 Db 2117 ACAGTCATGGCTACAGGAAATGATGATGCTTTAGAACCTCAAGATGAAGAGTGTAT 2176
 Qy 400 AlaThrTyrAsp 403
 Db 2177 GCTGACTATGAT 2188
 RESULT 12
 US-09-880-107-2275
 : Sequence 2275, Application US/09880107
 : Patent No. US20020142981A1
 : GENERAL INFORMATION:
 : APPLICANT: Horne, Darci T.
 : APPLICANT: Vockley, Joseph G.
 : APPLICANT: Scherf, Uwe
 : APPLICANT: Gene Logic, Inc.
 : TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 : FILE REFERENCE: 44921-5028-WO
 : CURRENT APPLICATION NUMBER: US/09/880,107
 : PRIOR FILING DATE: 2001-06-14
 : PRIOR APPLICATION NUMBER: US 60/211,379
 : PRIOR FILING DATE: 2000-06-14
 : PRIOR APPLICATION NUMBER: US 60/237,054
 : NUMBER OF SEQ ID NOS: 3950
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 2275
 : LENGTH: 6909
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : OTHER INFORMATION: Genbank Accession No. US20020142981A1 M16967
 US-09-880-107-2275
 Alignment Scores:
 Pred. No.: 0.00315 Length: 6909
 Score: 128.50 Matches: 85
 Percent Similarity: 33.02% Conservative: 55
 Best Local Similarity: 20.05% Mismatches: 145
 Query Match: 4.89% Indels: 139
 DB: 10 Gaps: 21
 US-10-088-045-2 (1-502) x US-09-880-107-2275 (1-6909)
 Qy 63 Aspa1a1leValThrh1sa1aPProgluValPProProValaAspArg----- 78
 Db 1162 GAAGTCATTGGAGCTATGCACTGTATATACCAAGATATGCAAAATAATACAGTCT 1221
 Qy 79 -----Asph1sProalalyValaValaValys 87
 Db 1222 CAGCATTGGATATTTCTCAACCAATGGAAACATTATAGAAAGTTATATACACA 1281
 Qy 88 Met-----GluThyValGluValMetArgLeuAlaAsp 99
 Db 1282 CAGTACGAGATGATGCTTCACCAACATACAGTATCCCAATATGAAA---GAAGAT 1338

Qy	100	GIYValIGluYrGInPheTrpTrhPheIGluYrGluValProGIuGImetIleuYrAl	119
Db	1339	GGG-----ATTGGGCTCTATTATTCAGAGC	1365
Qy	120	ArgGluGlyAerThrIleGluValGInPheSerAsn-----HisProAspSerLys	136
Db	1366	CAGGTGAGAGACACACATCCAAATCGTTCACAAATAATATGATGCGCGGCCCTCATATGACATT	1422
Qy	137	MetProHisAsnValAerPheHisValaIaThrGlyProGIuYrGlyValaGluAAser	156
Db	1426	TACCTTCATGAGATGACCTTCCTGCGCTTATGAAGATGA-----GTCAACTCTCT	1476
Qy	157	PheThrAlaProGIuYrHisThrSerThrPheSerPheValaIleuGInProGly---Leu	175
Db	1477	TTCACTCA---GGCAGAAACAACACATAG---ATCAGAGCAGTTCAACACAGGGGAAACC	1530
Qy	176	TyrValTyr-----HisCys	180
Db	1531	TATACTTATTAAGTGAACAACCTTAGAGTTTGATGAACCCACAGAAATATGATCCAGTGC	1590
Qy	181	AlaValAlaPro-----ValGluMetHisIleAlaAsnGlyMetTyr	194
Db	1591	TTAAACAAGACCATCTACTACAGTACGACCTGGACATCATGAGACATGCGCTTGCGCTAATA	1650
Qy	195	GlyLeuIleLeuValGInProLysGluGlyLeuProLys-----Val	208
Db	1651	GGACTACTCTTATCTGTGAAGACAGATCCCTGACAGCGGAGAAATACAGAGGGGACGA	1710
Qy	209	AerLysGluTyrTyrValMetGInGluAerPhe---TyrThrLysGlyLysTyrGlyGlu	227
Db	1711	GACATCGAAACAGACAGCGCTGTGTTCTGCTGTGTATGAGAAACAAAGCTGTACTCTTGA	1770
Qy	228	GInGlyLeuGInProPhe-----AerMetGluValaIleArgGluAspAlaGluTyr	245
Db	1771	GACAACATCAACAAGTTTGTGAATACTGTATGAGAGTGAAGACATGATGCCCAAGTTT	1830
Qy	246	ValValPheAsnGlySerValGlyValaLeuThrGluGluAsnAlaLeuValaLysVal	265
Db	1831	-----TATGATCAAAATC	1845
Qy	266	GlyGluThrValArgLeuPheValGlyuAsnGlyGlyProAsnLeuThrSerPheHis	285
Db	1846	ATGACACATCATCATAGGCTATGTG-----CTGAGAGCATTAAT	1884
Qy	286	ValIleGlyGluIlePheAsp-----	292
Db	1885	ACTCTTGATTTCTGTTGATGACACTGCTCCAGTGGCACTCTGTAGTGGGAGCCAG	1944
Qy	293	-----LysValHisPheGluGlyLys-----GlyGluAsnHisAsn	305
Db	1945	AATGAATAATTGACATCATCTTCACTGGGACATCTATCTATGGAAGAGGCATGAG	2004
Qy	306	IleGlnThrThrLeuIleProAlaGlyGlyValaAlaIleThrGlnPheLysValaAspVal	325
Db	2005	GACACCTTGACCCCTCTTCCCATGGCGTGAAGATCTGAGC-----	2046
Qy	326	ProGluAerTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyValaLeu	345
Db	2047	-----GTCACATGAT-----AATGTTGGAACCTTGG	2073
Qy	346	GlyIleLeuLysValaGluGlyGluGluAsnHisGluIleTyrSerHisLysGInThrAsp	365
Db	2074	ATGTTAACTTCCATGAATTCTATGTCGAAGACAAAGACGTAAGCTGAAATTCAGAGAT	2133
Qy	366	AlaValTyrLeuProGluGlyAlaProGlnAlaIleAerThrGlnGluAlaProLysTrh	385
Db	2134	GTTAAATGTATCCAGATGATGATGAAGACATCATATGATGATTTTGAACCTCCAGAAATCT	2193
Qy	386	ProAlaProAla-----AsnLeuGInGluGInIleLysValaGlyLys	399
Db	2194	ACAATCATGGCTAACAAGAAATGCATGATGCTTTAGAACTCTTAAGTGAAGACATGAT	2253
Qy	400	AlaThrTyrAsp	403

```

Db              2254 GCTGACTATGAT 2265

RESULT 13
US-10-115-563-13
: Sequence 13, Application US/10115563
: Publication No. US20030008307A1
: GENERAL INFORMATION:
: APPLICANT: Griffn, John H
:      Greengard, Judith S
: TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
:      C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
:      AND COMPOSITIONS THEREOF
: NUMBER OF SEQUENCES: 28
: CORRESPONDENCE ADDRESS: *
: ADDRESSER: The Scripps Research Institute, Office of
:      Patent Counsel
: STREET: 10666 No. US20030008307A1ch Torrey Pines Road, TPC 8
: CITY: La Jolla
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/115,563
: FILING DATE: 02-Apr-2002
: CLASSIFICATION: <unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/410,488
: FILING DATE: 24-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Fitting, Thomas
: REGISTRATION NUMBER: 34,163
: REFERENCE/DOCKET NUMBER: 449.0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-554-2937
: TELEFAX: 619-554-6312
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 6925 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: cDNA
:   FEATURE:
:     NAME/KEY: mat_peptide
:     LOCATION: 175..6765
:     OTHER INFORMATION: /product= "Factor V"
:   FEATURE:
:     NAME/KEY: sig_peptide
:     LOCATION: 91..174
:   FEATURE:
:     NAME/KEY: CDS
:     LOCATION: 91..6765
:   FEATURE:
:     NAME/KEY: misc_feature
:     LOCATION: 1..8
:   FEATURE:
:     NAME/KEY: misc_feature
:     LOCATION: 6918..6925
:   OTHER INFORMATION: /standard_name= "EcoRI linker
:     nucleotide sequence"
: SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-115-563-13

Alignment Scores:
Pred. No.:          0 00316      length:      6925
Score:             128.50      Matches:      85
Percent Similarity: 33.02%      Conservative: 55
  
```

Best Local Similarity: 20.05% Mismatches: 145
 Query Match: 4.89% Indels: 139
 DB: 14 Gaps: 21

US-10-088-045-2 (1-502) x US-10-115-563-13 (1-6925)

```

QY 63 ASPAIIIEValThrhIsaIAPROGluValProProProValAspArg-----78
DB 1162 GAAGTATTTGGAGCATGATGACCTGTATATACAGCGAATATGACAAATAATACAGGCT 1221
QY 79 -----AspHisProAlaLysValValLys 87
DB 1222 CAGCATTTGGATATTCTTCAAAACAAATGAGAAAAGTTATAGAAAGTTATGACACA 1281
QY 88 Met-----GluThrValGluLysValMetArgLeuAlaAsp 99
DB 1282 CAGTACGAAGATAGTCCTTACCAAAACATACAGTGAATCCCAATATGAAA---GAAAGT 1338
QY 100 GlyValGluTyrGlnPheThrPthrPheGlyGlyGlnValProGlyGlnMetIleArgVal 119
DB 1339 GCG-----ATTGGGTCTTATATACAGGCC 1365
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136
DB 1366 CAGGTACAGACACACTCAAAATCGTTCAAAAATATGCGCAGCCGCTTATAGCATT 1425
QY 137 MetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyGlnValGlnValAsp 156
DB 1426 TACCTTCATGAGATGACCTTCTGCTTATGAAAGTAA-----GTCAACTCTTCT 1476
QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu 175
DB 1477 TTCACCTCA---GGCAGAAACAAACACCATG---ATCAGACGACATTCAACAGGGAAC 1530
QY 176 TyrValTyr-----HisCys 180
DB 1531 TATACTTATATAGTGAACATCTTAGAGTTGATGAACCCACAGAAAATGATGCCCATG 1590
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
DB 1591 TTAACAAGACCATACTACAGTCGTCGATGATGAGACATCGCTCTGGGCTAATA 1650
QY 195 GlyLeuIleLeuValGluProLysGlyLeuProLys-----Val 208
DB 1651 GGACTCTCTTATCTGTAGAGACAGATCCCTGACAGCGGAGGAATACAGAGGACGA 1710
QY 209 AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyGlu 227
DB 1711 GACATCGAAGCAGCAGCGCTGTGTGCTGTGTGATGAGAAACAAAGCTGTGACTTGAG 1770
QY 228 GlnGlyLeuGlnProPhe-----AspMetGlnLysAlaIleArgLysAspAlaGluTyr 245
DB 1771 GACACATCAACAAAGTTTGTGAAAATCCTGATGAGGTGAAGAGTATGACCCCAAGTTT 1830
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyLysAsnAlaLeuLysAlaLysVal 265
DB 1831 -----TATGATCAAAACATC 1845
QY 266 GlyGluThrValArgLeuPheValGlyLysAsnGlyProAsnLeuThrSerSerPheHis 285
DB 1846 ATGAGCATATCAATGCTATGATG---CCTGAGACATTAATC 1884
QY 286 ValIleGlyGluIlePheAsp-----292
DB 1885 ACTCTTGATTTCTGCTTGTATGACACTGCCAGTGCACCTTCTGATGTGGGAGCCAG 1944
QY 293 -----LysValHisPheGlnGlyLys-----GlyGluAsnHisAsn 305
DB 1945 AATGAATTTTTCATCATCACTTCACTGCGCACTCATTCATATGAAAGAGCAGCATGAG 2004
QY 306 IleGlnThrThrLeuIleProAlaGlyGlyAlaIleThrGlnPheLysValAspVal 325
DB 2005 GACACCTGACCTTCTTCCCATGCGTGAGATCTGTGACG-----2046
  
```

```

QY 326 ProGlyAspTyrValLeuValAspHisAlaIlePheArgAlaPheAsnLysGlyAlaLeu 345
DB 2047 -----CTCAACATGAT-----AATGTTGGAACCTGG 2073
QY 346 GlyIleLeuLysValGluGlyGlnGluValAsnHisGlnIleTyrSerHisLysGlnThrAsp 365
DB 2074 ATGTTAACTTCCATGATTTCTAGTCCAAAGAACAAAGCTGAGCTGAATTCAGGGAT 2133
QY 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGlnValaProLysThr 385
DB 2134 GTTAAATGTTCCCAATGATGATGAGAGACTCATATGAGATTTTGAACCTCCAGAAATCT 2193
QY 386 ProAlaProAla-----AsnLeuGlnGluGlnIleLysAlaGlyLys 399
DB 2194 ACAGTCATGCTTACAGGAAATATGATGATCGTTTGAACCTGAAAGTGAAGAGATGAT 2253
QY 400 AlaThrTyrAsp 403
DB 2254 GCTGACTATGAT 2265
  
```

RESULT 14

US-10-115-563-26

Sequence 26, Application US/10115563

Publication No. US20030008307A1

GENERAL INFORMATION:

APPLICANT: Griffith, John H

TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN

C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION

AND COMPOSITIONS THEREOF

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSER: The Scripps Research Institute, Office of

Patent Counsel

STREET: 10666 No. US20030008307A1ch Torrey Pines Road, TPC 8

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/115,563

FILING DATE: 02-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/410,488

FILING DATE: 24-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: 449..0

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-554-2937

TELEFAX: 619-554-6312

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 6925 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 26:

US-10-115-563-26

Alignment Scores:

Pred. No.: 0.00316 Length: 6925

Score: 128.50 Matches: 85

Percent Similarity: 33.02% Conservative: 55

Best Local Similarity: 20.05%
 Query Match: 4.89%
 DB: 14
 Mismatches: 145
 Indels: 139
 Gaps: 21

US-10-088-045-2 (1-502) x US-10-115-563-26 (1-6925)

```

QY 63 AspaAlIleValThrHisAlaProGluValProProValAspArg----- 78
DB 1162 GAAGTATTGGAGCTATGCACTGTATATCCAGCGAATATGGACAAATAATACAGGTCT 1221
QY 79 -----AspHisProAlaLysValValLys 87
DB 1222 CAGCATTTGGATATTCTCAACCAATTGAAAACTTATAGAAAGTTATGTACACA 1281
QY 88 Met-----GluThrValGluLysValMetArgLeuAlaAsp 99
DB 1282 CAGTACGAAGATGAGTCTTCACCAACATACAGTGAATCCCAATATGAAA---GAGAT 1338
QY 100 GlyValGluLysGlnPheThrPheGlyGlnValProGlyGlnMetIleArgVal 119
DB 1339 GGG-----ATTGGGTCTTATTCAGAGCC 1365
QY 120 ArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136
DB 1366 CAGGTACAGACACACTCAAAATGCTTCAAAAATATGGCCAGCGCCCTATAGCATT 1425
QY 137 MetProHisAsnValAspPheHisAlaAlaThrGlyProGlyGlyValGluAlaSer 156
DB 1426 TACCTCATGAGTACCTTCGCGCTATGAAATGAA-----GTCAACTCTTCT 1476
QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGly---Leu 175
DB 1477 TTCACCTCA---GGCAGAACACACCATG---ATCAGACAGTTCAACAGGGGAAACC 1530
QY 176 TyrValTyr-----HisCys 180
DB 1531 TATACTTATTAAGTGAACATCTTAGAGTTTATGTAACCCACAGAAAATGATGCCAGTGC 1590
QY 181 AlaValAlaPro-----ValGlyMetHisIleAlaAsnGlyMetTyr 194
DB 1591 TTAACAGACCATACTACAGTACGTCGATGACATGAGACATCGCCTCGGCTATATA 1650
QY 195 GlyLeuIleLeuValGluProLysGluLysLeuProLys-----Val 208
DB 1651 GGACTACTTCTTAATCTGTAAAGACAGATCCCTGACAGGACAGAAATACAGAGGCGACA 1710
QY 209 AspLysGluTyrTyrValMetGlnGlyAspPhe---TyrThrLysGlyLysTyrGlyGlu 227
DB 1711 GACATCGAACACAGAGGCTGTGCTGTGCTGTGATGAGAACAAAGCTGTACCTTGAG 1770
QY 228 GlnGlyLeuGlnProPhe-----AspMetGluLysAlaIleArgGluAspAlaGluTyr 245
DB 1771 GACACATCAACAAAGTTTGTGAAATCTGTAGAGGTGAACGTGATGACCCCAAGTTT 1830
QY 246 ValValPheAsnGlySerValGlyAlaLeuThrGlyLysAlaLeuLysAlaLysVal 265
DB 1831 -----TATGATCAAAACATC 1845
QY 266 GlyGluThrValArgLeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHis 285
DB 1846 ATGAGCACTATCAATGGCTATGTG-----CCTGAGAGCATTAAT 1884
QY 286 ValIleGlyGluLysPheAsp----- 292
    
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 5, 2003, 10:29:59 / Search time 3065 Seconds
(without alignments)
3980.700 Million cell updates/sec

Title: US-10-088-045-2
Perfect score: 2626
Sequence: 1 MSKPTLITKTLICALSALML.....NKGQLSADVDVAKAKTKPN 502

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10088045/runat_04092003_083142_3928/app_query.fasta_1.647
-DB=BST -QFMT=fastap -SUFFIX=fst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10088045 @CGN 1.1 2810 @runat_04092003_083142_3928 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlmv:*
5: em_estlpl:*
6: em_estlro:*
7: em_hlc:*
8: em_hlc1:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_p2n:*
20: em_gss_vrt:*
21: em_gss_fnt:*
22: em_gss_mus:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_pro:*
26: em_gss_png:*
27: em_gss_vrt:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	624	23.8	554	28	BH380443
2	581.5	22.1	513	28	BH401929
3	542	20.6	461	28	BH395696
4	303	11.5	263	28	BH375921
5	264.5	10.1	796	28	BH404724
6	210	8.0	795	28	BH380839
7	186.5	7.1	653	12	B1180641
8	182	6.9	719	29	B2337383
9	170.5	6.5	353	28	BH380607
10	158	6.0	492	29	B2716485
11	157	6.0	754	10	BG646015
12	155.5	5.9	700	28	BH378995
13	154.5	5.9	814	29	B2574542
14	152.5	5.8	413	9	AL377723
15	146	5.6	612	14	CB882241
16	145.5	5.5	501	14	CA553053
17	145.5	5.5	513	14	CA554730
18	145.5	5.5	515	14	CA551823
19	145.5	5.5	515	14	CA552719
20	145.5	5.5	1145	29	B2559318
21	144.5	5.5	567	10	BF648894
22	144.5	5.5	596	29	CC016516
23	143.5	5.5	360	14	CA550887
24	143.5	5.5	607	12	BH318801
25	142.5	5.4	623	9	AV822649
26	142.5	5.4	691	14	CA483187
27	142	5.4	752	28	AY079863
28	141.5	5.4	502	14	CA554547
29	141.5	5.4	600	9	AV826062
30	141.5	5.4	1269	29	B2560121
31	141	5.4	770	29	B2535086
32	140.5	5.4	615	14	CA553237
33	139.5	5.3	614	9	AV826269
34	139.5	5.3	659	10	BG448142
35	139.5	5.3	690	10	BF644576
36	139	5.3	571	9	AV826844
37	138.5	5.3	489	14	CA553614
38	138	5.3	645	10	BF650165
39	137	5.2	737	13	BU635352
40	136.5	5.2	612	9	AV826343
41	136.5	5.2	942	29	CNS070GT
42	134.5	5.1	563	12	B0226252
43	133.5	5.1	523	9	AW695071
44	133	5.1	620	10	BE920041
45	133	5.1	650	10	BG135584

ALIGNMENTS

RESULT 1
BH380443
LOCUS BH380443
DEFINITION AG-ND-168117.TR ND-TAM Anopheles gambiae genomic clone AG-ND-168117
ACCESSION BH380443
VERSION BH380443.1 GI:17326585
KEYWORDS
SOURCE GSS.
ORGANISM Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.

Query Match: 22.14% Indels: 1
DB: 28 Gaps: 1
US-10-088-045-2 (1-502) x BH401929 (1-513)

QY 117 ILEAGVALARGGLUGLYSPHTRIIIEGLUVALGlnPheSerAsnHisProAspSerLys 136
DB 7 ATCCGGGCAAGAGTGGGTACGCAAAATTCATTGGAATGATTAATAAACAAGTACA 66
QY 137 MetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluLaser 156
DB 67 TTCGCCGATATATATACCTTCACTGCTTAATGCGCTGGTGGCGGTGACAGAGTACT 126
QY 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyr 176
DB 127 TTGTAGCTCCGGGAAGAGAGCGGATTTAACTTCAAAAGCCCTTAATCCAGGTCTGTC 186
QY 177 ValTyrHisCysAlaValAlaProValGlyMetHisAlaAsnGlyMetTyrGlyLeu 196
DB 187 GATATCCACTGTGCACCGCTCCGGTGGGAATGCATATTCCTAATGGATGATGATTA 246
QY 197 ILEAGVALARGGLUGLYSPHTRIIIEGLUVALAspLysGlyTyrTyrValMetGln 216
DB 247 ATTCGATATGACCGGAGAGTGAATTCCTTAAGATGATTAAGATTCATATCATGCG 306
QY 217 GlyAspPheTyrThrLysGlyLysTyrGlyGlyGlnGlyLeuGlnProPheAspMetGlu 236
DB 307 GGAGATTTCTATACCAAGCTAAATACGAGACAAAGCTCTTCAGGAATTTGATATGAT 366
QY 237 LysAlaIleArgGlyAspAlaGlyTyrValAlaPheAsnGlySerValGlyAlaLeuThr 256
DB 367 AAGCAATTTGACCAATCTCAATATATGCTTTTCAATGATTAACAGACACACTTCG 426
QY 257 GlyIleuAsnAlaLeuLysAlaLysValGlyIleuThrValArgLeuPheVal 273
DB 427 ---GATATGCACTGAAGAAGGTTGGTGAATAATGTAAGATTTCTTGT 474

RESULT 3
BH395696 461 bp DNA linear GSS 11-DEC-2001
LOCUS AG-ND-105E5.TF ND-TAM Anopheles gambiae genomic clone AG-ND-105E5,
DEFINITION BH395696
VERSION BH395696.1 GI:17341837
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
REFERENCE 1 (bases 1 to 461)
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Hille, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.
Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
JOURNAL MEDLINE PUBLISHED
COMMENT Other GSSs: AG-ND-105E5.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PBST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Classes: BAC ends.
FEATURES
source location/Qualifiers
1..461
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PST"
/db_xref="taxon:7165"
/clone="AG-ND-105E5"
/clone_1ib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 148 a 71 c 116 g 126 t
ORIGIN
Alignment Scores:
Pred. No.: 1,13e-46 Length: 461
Score: 542.00 Matches: 107
Percent Similarity: 79.74% Conservative: 15
Best Local Similarity: 69.93% Mismatches: 30
Query Match: 20.64% Indels: 3
DB: 28 Gaps: 0

US-10-088-045-2 (1-502) x BH395696 (1-461)

QY 116 MetIleArgValArgGlyGlyAspThrIIIEGLUVALGlnPheSerAsnHisProAspSer 135
DB 3 CTTATATAGCGCAAGAGTGGGTGACCAATTCATTTTCAGAAATATGATAAACAAGT 62
QY 136 LysMetProHisAsnValAspPheHisAlaIleThrGlyProGlyGlyValAlaGluLys 155
DB 63 ACATTCGCCGATATATATGACCTTCATCTGTAATGCGCTGGTGGCGGTGACAGAGCT 122
QY 156 SerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeu 175
DB 123 ACTTTTGTAGGTCGGGATTTT--GCGGTATTTTAACTTCAAAAGCCCTTAATCCAGGTCTG 181
QY 176 TyrValTyrHisCysAlaValAlaProValGlyMetHisAlaAsnGlyMetTyrGly 195
DB 182 TACGATATCCACTGTGCACCGCTCCGGTGGGAATGATATTCCTAATGCGATGATGGA 241
QY 196 IleuIleuValGluProLysGlyLysLeuProLysValAspLysGlyTyrTyrValMet 215
DB 242 TTAATTCGATATGAAACCGAAGTGGATTCCTTAAGATGATTAAGAGTCTTATATCATG 301
QY 216 GlnGlyAspPhe-TyrThrLysGlyLysTyrGlyGlyGlnGlyLeuGlnProPheAspMe 235
DB 302 CAGGAGATTTCAATATACCAAGGTAATACGAGACAAAGGCTTCAGGAATTTGATAT 361
QY 235 tGluIleAlaIleArgGlyAspAlaGlyTyrValAlaPheAsnGlySerValGlyAlaLe 255
DB 362 GATATTAAGCAATTCGAGAACTCCGATATATGCTTTTCAATGATTAAGCAGAGACT 421
QY 255 tTyrGlyLysAsnAlaLeuLysAlaLysValGlyGlu 267
DB 422 TCTGGGAGATATGAACTGAAGA-AAGGTTGGTGA 457

RESULT 4
BH375921 263 bp DNA linear GSS 10-DEC-2001
LOCUS AG-ND-119L15.TF ND-TAM Anopheles gambiae genomic clone AG-ND-119L15,
DEFINITION BH375921
VERSION BH375921.1 GI:17322063
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
REFERENCE 1 (bases 1 to 263)
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Hille, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.
Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
JOURNAL MEDLINE PUBLISHED
COMMENT Other GSSs: AG-ND-105E5.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org

REFERENCE AUTHORS

1 (bases 2 to 263)
Hong, Y.-S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carille, J.L., Black, K., Zhang, H.-B., Gardner, M.J. and Collins, F.H.
Construction of a BAC library and generation of BAC end

TITLES

sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*
Mol. Genet. Genomics 268 (6), 720-728 (2003)
02542063

PUBMED
COMMENT

12655398
Other_GSSs: AG-ND-119L15.TR

TITLE

sequence-tagged connectors for genome sequencing of the African malaria mosquito <i>Anopheles gambiae</i>	
Mol. Genet. Genomics 268 (6), 720-728 (2003)	
JOURNAL MEDLINE	22542063
PUBMED	12655398
Other GSSs:	AG-ND-139F12.TF.1
COMMENT	

COMMENT

the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: MJ3 For
class: BAC ends.

FEATURES

Location/Qualifiers
1. .263

```
/organism="Anopheles gambiae"  
/mol_type="genomic DNA"
```

```
/strain="PEST"  
/db_xref="taxon:7165"
```

```
/clone="AG-ND-119L15"  
/clone_lib="ND-TAM"
```

BASE COUNT
ORIGIN

76 a 48 c 67 g 72 t

Alignment Scores:

Pred. No.:	6,23e-22	length:	26
Score:	303.00	Matches:	57
Percent Similarity:	75.29%	Conservative:	7
Best Local Similarity:	67.06%	Mismatches:	20
Query Match:	11.54%	Indels:	1
DB:	28	Gaps:	0

US-10-088-045-2 (1-502) X BH375921 (1-263)

OY	11	IIeArgValArgGluIuIyAspRhrIzIegValAlaInPheSerAenH:PtoAsnSerLys	134
Db	11	ATCAGGGCAAGAGTGGGTGCACGAATTGAAATTTCATTTTGAGMATAATGAAAACAGTACA	70
OY	137	MeProHisAenValAspPheHisAlaAlaathrGlyProGlyIyGlyValAlaGlnAlaSer	156
Db	71	TTCGCCCATATATATACCTTCATCTGTAATAGCCCTGGGGGGCGCAAGACT	130
OY	157	PheThrAlaProGlyHisIsthiserThrPheSerPheIyValAlaGlnProGlyLeuTyrr	176
Db	131	TTTTGATGCTCCGGGAAGAACGCGGTATTTAACTTCGAAGCCCTCAAAATCCACAGCTCTAC	180
OY	177	ValTYTHISCyAlaIValAlaLaProValGIymethisIleAlaasngIymetyr-GlyLe	196
Db	191	GATACCAATGTCGACACCGCTTCCCGTGGGATGATCATTTGCTAATGGGATGATATGGGATT	250
OY	196	IuIeLeuValGlu	200
Db	251	AATTCTGATGAGA	263

LOCUS
DEFINITION

LOCUS	BH404724	796 bp	DNA	linear	GSS 11-DEC-2000
DEFINITION	AG-ND-139F12.TR.1 ND-TAM Anopheles gambiae genomic clone				
ACCESSION	AG-ND-139F12, genomic survey sequence.				
	BH404724				

ACCESSION
WEEDSTON

VERSION	BH404/24.1	GI:17350940
KEYWORDS	GSS.	
SOURCE	Amphibolite	15-01-2017

SOURCE

ORGANISM

ORGANISMS

Euarchyotia; metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocestera; Culicoidea; Anopheles.

REFERENCE

AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carille, J.L., Black, K., Zhang, H.-B., Gardner, M.J. and Collins, F.H.

RESULTS

sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*
Mol. Genet. Genomics 268 (6), 720-728 (2003)

JOURNAL
MEDLINE 22542063
PUBMED 12655398

Other GSSs: AG-ND-139P12.TF.1

COMMENT

COMMENT

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimise the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 Rev
Class: BAC ends.

FEATURES

source	1.	.796
--------	----	------

```
/organism="Anopheles gambiae"
/mol_type="genomic DNA"

```

```

/strain="PES1"
/db_xref="taxon:7165"

```

```

/clone="AG-ND-139F12"
/clone_lib="ND-TAM"
/notes="Notes"

```

BASE COUNT

ORIGIN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

Alignment Scores:

Printed No.:	3,120-17	Length:	79
Score:	264.50	Matches:	55
Percent Similarity:	52.868	Conservative:	19
Best Local Similarity:	39.294	Mismatches:	49
Query Match:	10.074	Indels:	17
DB:	28	Gaps:	3

03-10-088-045-2 (1-502) X BH404724 (1-796)

```

QY      11 leuilecyrsalaleuSerAlaIleuMet---leuSerGIYcySerSerangInAlaAspLys 29
Db      425 CTTCGCTGTTGTTTGTCTGCAATTTTACCTTGAAGCTCTGTAAACAAACGCTTCGAA 488
QY      30 AlaAlaGlnProLysSerSerThrValAlaPhaIaIaIaIaIaIaIaIaAspAsn 49
Db      485 GCTTCACAA-----TCCGAGAGTACGTCMAAGCTTTCTCTGTACGGAAATTCGAAG-- 538
QY      50 AlaAlaSerGlnGluIaIaGlnGlyIuLeuProValIleAspAlaIaIeValIThrIaIa 69
Db      536 -----GAGCGCAACATGACACTTCACCT 556

```


/lab host="DH10B (phage-resistant)"
 /clone.lib="LRH (lin rhodamine and Hoechst dye)"
 /note="Vector: pZLI, Site 1: SalI site; Site 2: EagI site;
 LRH library was constructed from cDNA of primary bone
 marrow cells depleted of lineage-committed cells and
 enriched for primitive cells by FACS sorting for cells
 with low level staining with rhodamine123 and Hoechst
 33342 dyes. cDNA from 5000 cells derived from 30 mice,
 5000 cells were were directionally cloned into SalI-EagI
 restriction site of the (Ziplox (Gibco BRL). The ligated
 cDNA fragments were transformed into DH10B host cells.
 The original library had an initial plating complexity of
 1.44X10⁷ clones."

BASE COUNT 118 a 200 c 208 g 125 t 2 others

ALIGNMENT SCORES:

Pred. No.: 3,63e-09 Length: 653
 Score: 186.50 Matches: 53
 Percent Similarity: 48.25% Conservative: 16
 Best Local Similarity: 37.06% Mismatches: 59
 Query Match: 7.10% Indels: 15
 DB: 12 Gaps: 6

US-10-088-045-2 (1-502) x B1180641 (1-653)

QY 366 AlaValTyrLeuProGluGlyAlaProGlnAlaIleAspThrGlnGluAlaProLysThr 385
 DB 613 GCCCGCTACCTCAAGACCTTCCGCCG-----GTGATCCGAGNAGCGGCTCAGCC 560
 QY 386 ProAlaProAlaAsnLeuGln-----GluGlnIleLysAlaGlyLysAla 400
 DB 559 TATGACCGCGAGCGGCGACAGCCCTGTGGAAGGATGACAGCAAGCGGGGGCTGC 500
 QY 401 ThrTyrAspSerGlnCysAlaAlaCysHisGlnProAspGlyLysGlyValProAsnAla 420
 DB 499 GTCTATATCGACACACTGCGCGCTGCGACCGTACCGCTATACCGCGGCTG 440
 QY 421 PheProProLeuAlaAsnSerAspTyrLeu---AsnAlaAspHisAlaArgAlaAsp 439
 DB 439 TTCCCGCGCTGGCCGGTATCCGGTGTGCAATCGCGGACCCGACGCTTGATCAT 380
 QY 440 IleValAlaAsnGly-----LeuSerGlyLysIleThrValAsnGlyAsnGlnTyrGlu 457
 DB 379 ACCGTCTCAAGCGCGGACCTTGCAGCGGACCAACCGCAGCTTCGAAAC----- 329
 QY 458 SerValMetProAlaIleAla-----LeuSerAspGlnGlnIleAlaAsnValIleThr 475
 DB 328 TTCACCATGCGCGCTTTCCTGCGATGTCGACAGGAAGTGGCGGACGTGCGGT 269
 QY 476 TyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAspAspValAlaLys 495
 DB 268 TTCATTCGCGGCGGCGGTGGGTATATCGCGCGGCGGACGCAAGCGCATGTGGCGG 209
 QY 496 AlaLysLys 498
 DB 208 TTGCGCAAG 200

RESULT 8 BZ337383 719 bp DNA linear GSS 06-NOV-2002
 LOCUS BZ337383
 DEFINITION 1a87a06.g1 WGS-Sbicolor (JM107 adapted methyl filtered) Sorghum
 bicolor genomic clone 1a87a06 5', genomic survey sequence.

ACCESSION BZ337383.1 GI:24731977
 VERSION BZ337383.1
 KEYWORDS GSS.
 SOURCE Sorghum bicolor (sorghum)
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 719)
 AUTHORS Rabinowitz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,

TITLE
JOURNAL
COMMENT

Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
 Zlatavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.
 Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
 Unpublished
 Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884
 Fax: 516 367 8874
 Email: mcombie@csghl.org
 Plate: 1a87 row: a column: 06
 Seq primer: -21M3univrev
 Class: shotgun
 High quality sequence stop: 719.
 Location/Qualifiers

FEATURES
source

1..719
 /organism="Sorghum bicolor"
 /mol_type="genomic DNA"
 /db_xref="taxon:4558"
 /clone="1a87a06"
 /lab_host="JM107 or DH5a"
 /note="Site 1: Xba I; Site 2: Xba I; The vector was filled in
 digested with Xba I and one nucleotide was added by fill in
 in the recessive 3' end. The genomic DNA was nebulized,
 end repaired, adaptor ligated and size fractionated using
 sephadex. The resulting fragments were between 0.8 and 3
 kb and were cloned into the vector (x/y reads in M13mp19,
 b/g reads in pUC19). The same ligation was transformed in
 either JM107 or DH5a."

BASE COUNT

148 a 232 c 214 g 125 t

ALIGNMENT SCORES:

Pred. No.: 1.23e-08 Length: 719
 Score: 182.00 Matches: 44
 Percent Similarity: 55.14% Conservative: 15
 Best Local Similarity: 41.12% Mismatches: 42
 Query Match: 6.93% Indels: 6
 DB: 29 Gaps: 4

US-10-088-045-2 (1-502) x BZ337383 (1-719)

QY 396 LysAlaGlyLysAlaThrTyrAspSerAsnCysAlaAlaCysHisGlnProAspGlyLys 415
 DB 42 AAGCGGCGCGCTGCTGTATATCGACACTGCGCGCTTGCACCGCAGCGGCAC 101
 QY 416 GlyValProAsnAlaPheProProLeuAlaAsnSerAspTyrLeu---AsnAlaAspHis 434
 DB 102 GGTACACCGGTGATATCCCGCGCTGCGGCAAGCCGCTGTTCAGACGCGGATGCG 161
 QY 435 AlaArgAlaAspIleValAlaAsnGlyLeuSerGlyLysIleThrValAsnGlyAsn 454
 DB 162 ACGTCTTGATCATATATCTGTGAACGCG-----GCAACCTTCCGCGCACCCACAG 215
 QY 455 GlnTyrGluSer---ValMetProAlaIleAla-----LeuSerAspGlnGlnIleAla 471
 DB 216 GCGCGCTTACCTTACCATGATCCCGCGCTTTCCTGCGCGCTTTCGCGCAAGTGGCG 275
 QY 472 AsnValIleThrTyrThrLeuAsnSerPheGlyAsnLysGlyGlnLeuSerAlaAsp 491
 DB 276 GATGTGTCACTTTCGTCGCTGCGGCGGATGCGGATGCGGCGGCGGCGGCGGCG 335
 QY 492 AspValAlaLysAlaLysLys 498
 DB 336 GACGTGCGGCGGACCTGCGCAAG 356

RESULT 9 BH380607 353 bp DNA linear GSS 10-DEC-2001
 LOCUS BH380607/c
 DEFINITION AG-ND-164017.TF ND-TMW Anopheles gambiae genomic clone AG-ND-164017
 , genomic survey sequence.

QY 145 -----AlaAlaThrGlyProGlyGlyValAlaGluAspSerPheThrAlaPro 160
 DB 196 CCGGACGCCCGGCGGCGCCACGACGAGGAGGCGTACGACGACGCCCATC---CTCCCC 252
 QY 161 GYHsIstHsSerThPheSerPheLysAlaLeuGlnProGlyLeuValTyrosCys 180
 DB 253 GGGGACACCTTACCTACGCTTCGTGTGAGACCGCGCGGACGTACATGTACAC--- 309
 QY 181 AlAlaAlaProValGlyMethIstIleAlaSerGlyMethTyrGlyLeuValGlu 200
 DB 310 -----GGCCACTACCGGATGACGCGCTCCGCGGCGCTCAACGCGCTCATGTGTGCGC 363
 QY 201 ProLysGluGlyLeuPro-----LysValAspLysGluTyrTyrValMet 215
 DB 364 GGGGCG 423
 QY 216 GlnGlyAspPheThrThyGlyGlyTyr 225
 DB 424 CTCACGACTGTGTGATTAAGACACCTAC 453
 RESULT 11
 BG646015 754 bp mRNA linear EST 24-APR-2001
 LOCUS EST507634 KV3 Medicago truncatula cDNA clone pKV3-48A4 5' end, mRNA
 DEFINITION sequence.
 ACCESSION BG646015
 VERSION BG646015
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 REFERENCE 1 (bases 1 to 754)
 Vandenbosch, K., Endre, G., Hur, J., Beremand, P., Town, C.D., Bowman
 C.L., Craven, M.B., Cho, J., and Fraser, C.M.
 ESTs from roots of Medicago truncatula 72 h after Rhizobium
 inoculation, 2001
 JOURNAL Unpublished
 COMMENT Contact: Vandenbosch K
 Department of Plant Biology
 University of Minnesota
 220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
 Tel: 612 624 2755
 Fax: 612 625 1738
 Email: kvandenb@cds.umn.edu
 M394422e TIGR sequence name: MTECJ02TK More information is
 available at: www.medicago.org
 Seq primer: Skmod (CTA gAA CTA gtc gat CC).
 FEATURES
 source
 1..754
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="genotype AL7"
 /db_xref="taxon:3880"
 /clone="pKV3-48A4"
 /tissue_type="Seedling roots"
 /dev_stage="3 days post-inoculation with Sinorhizobium
 meliloti"
 /lab_host="E. coli strain XL0LR"
 /clone_1ib="KV3"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 StrataGene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XL0LR cells."

Pred. No.: 5,51e-06 Length: 754
 Score: 157.00 Matches: 72
 Percent Similarity: 38.70% Conservative: 29
 Best Local Similarity: 27.59% Mismatches: 100
 Query Match: 5.98% Indels: 60
 DB: 10 Gaps: 14
 US-10-088-045-2 (1-502) x BG646015 (1-754)
 QY 99 AspglyValGluTyrGlnPheThrPhePheGlyGlyGlnValProGlyGlnMetLearg 118
 DB 5 GATTGTAAAGACCAATGTGTGATGGAATGCAATGCCAATTTCCAGGCCCACTATTAGA 64
 QY 119 ValArgGluGlyAspThrIleGluValGlnPheSerAsn-----HisProAspSerLys 136
 DB 65 GCTGAAGTTGGTGAACACTCTTGTATTGACCTCACACACAGCTCTATACAGAGGAACT 124
 QY 137 MetProHisAsnValAspPheHisAlaAlaThrGlyProGly-----Gly 151
 DB 125 GTTATTTCAC-----TGGCATGGAATCAGACAGTTTGGAACTCTGGGCTGATGA 175
 QY 152 GlyAlaGluAlaSerPheThrAla-----ProGlyHisThrSerThPheSerPheLys 169
 DB 176 ACTGCTGCATATCTCAATGCTCTATTAATCCAGAGAAACTTTCAATCAAAATTGCAA 235
 QY 170 AlaLeuGlnProGlyLeuTyrValTyrHisCysAlaValAlaProValGlyMethIstIle 189
 DB 236 GTTGACAGGCTGTGTGATCATATTCTATCAT-----GGACACTATGTATGCCAAGA 286
 QY 190 AlaAsnGlyMetTyrGlyLeuIleLeuValGluProLysGluGlyLeuProLysValAsp 209
 DB 287 GCAGACGGGTGTATGTTGTTCTCTAATAGTGAT-----TTACCAAGAGCCAA 334
 QY 210 LysGluTyrTyrValMetGlnLysAspPhe-----TyrThr 221
 DB 335 ACAGAGCCATTTTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 394
 QY 222 LysGlyLysTyrGlyGluGlnGlyLeuGlnProPheAspMetGluLysAlaIleArgGlu 241
 DB 395 AGTTACATGAAACAAAGAGGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 451
 QY 242 AspAlaGluTyrValVal-----PheAsnGlySerValGlyAlaLeuThr 256
 DB 452 CCACAGAGCTGTCTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511
 QY 257 GlyLysAsnAlaLeu-----LysAlaLysValGlyGluThrValArgLeuPheVal 273
 DB 512 GGGAGCACCAACCTACCCCAATGCAATTTGAAAGGTGTGAAGAAATGT----- 559
 QY 274 GlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyGluIlePheAspLys 293
 DB 560 -----GCACCCCAAGTT----- 571
 QY 294 ValHisPheGluGlyGlyLysGlyLysAlaSerAsnIle-----GlnThrThrLeuIlePro 312
 DB 572 CTTCACGTGTGAGCAAAAAAGACCTATAGATATAGATATAGATATAGATATAGATATAGAT 631
 QY 313 AlaGlyGlyAlaAlaIleThrGluPheLys-----ValAspValProGlyAspTyr 329
 DB 632 TCTCTCACTGGCGCATTTCAATCAACAACTTATTGTAGTGAAGACAGATGAAACTAT 691
 QY 330 Val 330
 DB 692 GTG 694
 RESULT 12
 BH378995/c 700 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-162A8.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162A8,
 DEFINITION genomic survey sequence.
 ACCESSION BH378995
 VERSION BH378995.1 GI:17325137
 KEYWORDS GSS.

SOURCE
ORGANISM Anopheles gambiae (African malaria mosquito)

REFERENCE
AUTHORS 1 (bases 1 to 700)
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carille, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.
Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
12655398
JOURNAL MEDLINE PUBMED
COMMENT Other GSSs: AG-ND-162A8.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M3 For
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..700
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-162A8"
/clone_1lb="ND-TAM"
/note="vector: pECBAC1; Site_1: HindIII"

BASE COUNT 185 a 156 c 116 g 243 t
ORIGIN

Alignment Scores:
Pred. No.: 7,15e-06 Length: 700
Score: 155.50 Matches: 38
Percent Similarity: 50.94% Conservative: 16
Best Local Similarity: 35.85% Mismatches: 34
Query Match: 5.92% Indels: 18
DB: 28 Gaps: 3

US-10-088-045-2 (1-502) x BH378995 (1-700)

QY 11 LeuUcYsAlaLeuSerAlaLeuMet---LeuSerGlyCySerAsnGlnAlaAsp-Ly 29
DB 212 CTTCGCTGTTGTTGTCGCAATTTTCACCTTGAAAGCTTGTAACAAACCTTCCTGA 213
QY 29 sAlaAlaGlnProLysSerSerThrValAspAlaAlaAlaLysrThrAlaAsnAlaAspAs 49
DB 212 AGCTTCACAA-----TCCGGAAGTACTGAGGTATTACTGTTCAAGGAAATCGAGA-- 161
QY 49 nAlaAlaSerGlnGluHisGlnGlyGluLeuProValIleAspAlaIleValIleHisAl 69
DB 160 -----GAAGCCAAACTGACATCAC 141
QY 69 aProGluValProProValAspArgAspHisProAlaLysValValValValMetC 89
DB 140 TTCAGATGCGCTTGAACCAATGAGAAACAGAGCTGCAAAAAGCTGATGTGACGTCGGA 81
QY 89 uThrValGluLysValMetArgLeuAlaAspGlyValGluTyrGlnPheTrpThrPheG 109

DB 80 AACATATGAAAAAACAAGTGAATGTGACAGACCAACCAATATTAATCTTGACCTTTTG 21
QY 109 YGlyGlnValProGly 114
DB 20 CGAAACAGTTCGGGA 5

RESULT 13
BZ574542/c
LOCUS
DEFINITION msh2_3719.y2 msh Pseudomonas aeruginosa genomic clone msh2_3719,
genomic survey sequence.
BZ574542
ACCESSION BZ574542
VERSION BZ574542.1 GI:27209603
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 814)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
Whole-genome-sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: ckraymond@u.washington.edu
Class: shotgun.

FEATURES
source Location/Qualifiers
1..814
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_3719"
/clone_1lb="msh"
/note="Environmental isolate. Whole genomic shotgun library."

BASE COUNT 127 a 243 c 283 g 160 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1.12e-05 Length: 814
Score: 154.50 Matches: 53
Percent Similarity: 42.31% Conservative: 35
Best Local Similarity: 25.48% Mismatches: 79
Query Match: 5.88% Indels: 41
DB: 29 Gaps: 9

US-10-088-045-2 (1-502) x BZ574542 (1-814)

QY 325 ValProGlyValAspTyrValLeu-----ValAspHisAlaIlePheArgAlaPhe 340
DB 774 GTTCCGGGCAAGTTTACTCGACGCGCAACCGTCGAACAT-----TGGCAACCGCTG 721
QY 341 Asn-----LysGlyAlaLeuGlyIleLeuLysValGluGlyGlu 353
DB 720 AGCTTGCCCACTTGTGTGACGCGTGAACCCCGTCGACTTCTGAAGACC---GGCGAG 664
QY 354 GluAsnHisGluIleTyrSerHisIlePheGlnThrAspAlaValTyr----- 368
DB 663 ANCCGCTTCGACACGCGTCCGCGACATGACCGATGTCATCCACCACCAAGCACCGACAC 604
QY 369 -----LeuProGluGlyAla 373
DB 603 TTCAGCGACGACGATCTGCTGCGCATCGCCACGTAACCTGAAGTCCCTCCGCGCGCAAG 544
QY 374 ProGlnAlaIleAspThrGlnGluAlaProLysThrProAlaProAlaAsnLeuGlnGlu 393

Fax: 039482-5595
Email: stein@pk-gatersleben.de
Insert length: 612 Std Error: 0.00
Plate: 1 row: E column: 24
Seq primer: T7

FEATURES

source

Location/Qualifiers

1. 612
/organism="Hordelum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Daike"
/db_xref="GABI:557881"
/clone="HL01B24"
/issue_type="embryo, scutellum und aleuron"
/dev_stage="seeds, 12 h and 62 h after imbibition"
/lab_host="DH10B"
/clone_lib="HL"
/note="Vector: pSPORT1; Site_1: SalI (5'-end of cDNA); Site_2: NotI (3'-end of cDNA); average insert size is 1.6-1.7 kb, library was created by RZPD, Heidelberg (Dr. Bernhard Korn, Im Neuenheimer Feld 506, D-69120 Heidelberg Germany, http://www.rzpd.de)"

BASE COUNT 114 a 187 c 187 g 124 t
ORIGIN

Alignment Scores:

Pred. No.: 5.9e-05 Length: 612
Score: 146.00 Matches: 47
Percent Similarity: 42.67% Conservative: 17
Best Local Similarity: 31.33% Mismatches: 58
Query Match: 5.56% Indels: 28
Gaps: 7

US-10-088-045-2 (1-502) x CB882241 (1-612)

93 LysValMetArgLeuAlaAspGlyValGluTyrGlnPheTyrThr----- 107
168 AAGGCCAGGACCTCAAGTGGAGGAGGACCACTTCTGTCGCCGAGCTGCGAGGAG 227
108 -----PheGlyGlyGlnValProGlyGlnMetLeuArgValArgGluGly 122
228 AAGGCTTCATGGCATCAACGCCGACGTTCCCGGCCGCCACCATCCGGCCAGGCCGCG 287
123 AspThrIleGluValGlnPheSerAsn-----HisProAspSerIleMetProHisAsn 140
288 GACACCATGTGTGTGAGCTCAAGAACGGGCTGCACACCGAGGGGTGTCTTCAC--- 344
141 ValAspPheHisAlaAlaThrGlyProGly-----GlyGlyAlaGluAla 155
345 -----TGGCAGCGGCTCAGACAGATTGGACACCGTGGCGGATGGCAGCGCGCATC 398
156 SerPheThrAla-----ProGlyHisThrSerThrPheSerPheValAlaLeuGlnPro 173
399 TCCCATGCGGCATCAACCCCGAGAAACCTTCACTTACCGATTGTGTCTGCACAGACGG 458
174 GlyLeuTyrValTyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMet 193
459 GGGACATATTTTACCAT-----GGGCACTACGCGATGACAGAGCGCGGCGCTG 509
194 TyrGlyLeuIleLeuValGluProLysGluGlyLeuPro-----LysValAspLys 210
510 TACGCTCTCTGATGTGTGATGTGCAGATGGAGAGAACCCGTTCAAGTATGACGGC 569
211 GluTyrTyrValMetClnGlyAspPheTyr 220
570 GAGCTGAACCTGTCTCTCAGCAGCTGTAC 599

RESULT 16

CAS53053/3 501 bp mRNA linear EST 19-NOV-2002
LOCUS
DEFINITION C0866F05-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus
ACCESSION CDNA clone NIA:C0866F05 IMAGE:3001360 5', mRNA sequence.
CAS53053

VERSION CAS53053.1 GI:25097296
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 501)
AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)
JOURNAL Unpublished
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: C0866 row: F column: 05
Seq primer: M13 Reverse
High quality sequence stop: 501
POLYA=No.

FEATURES

source

Location/Qualifiers

1. 501

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="nlaEST:C0866F05-5N"

/db_xref="taxon:10090"

/clone="NIA:C0866F05 IMAGE:3001360"

/issue_type="Blastocyst"

/dev_stage="3.5-dpc"

/lab_host="DH10B"

/clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"

/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544191]). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo(dT) primer (Invitrogen): 5'-DCACTAGTTTGTAGATCGAGCGCGCCCTTTTCTTTT-3' from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker IL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT

98 a 152 c 146 g 105 t

ORIGIN

Alignment Scores:

Pred. No.: 5.06e-05 Length: 501
Score: 145.50 Matches: 44
Percent Similarity: 43.11% Conservative: 28
Best Local Similarity: 26.35% Mismatches: 66
Query Match: 5.54% Indels: 29
DB: 14 Gaps: 8

US-10-088-045-2 (1-502) x CAS53053 (1-501)

93 ProValIleAspAlaIleValThrHisAlaProGluValProProProValAlaAspArg 79
498 CCGATGACGACAG--GCGACGATGCGACCGCTTGCATCCACGCTCCGCTCCGAC 442
80 HisProAlaLysValValVal-----LysMetGlu----- 89

Db 441 TATCGGCGCGGTGTAACGCTGAACCGAGCTGTCGTTGCCATGCGCATGAACGCGCATG 382
 Oy 90 -----ThyValIGluysValMetArg---LeuAlaSerGlyValGluTyr 103
 Db 381 AAGGAGTTTCACTGTGGCCGCAACCTGTGGTGGCGAAATTGGCCGAAGCGCAAGAGGCT 322
 Oy 104 GlnPheTrpThrPheGlyGlyGlnValProGlyGlnMetIleArgValArgGluGlyAap 123
 Db 321 CATCTGGGGGTTTCAACGGCCAACTCCCGGGGCCACCATTTGAGGCAGTACAAAGCGAC 262
 Oy 124 ThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsn-----Val 141
 Db 261 AAGCTCCGATATTCGTGCACCAAC-----AAATTGCCGAGCATACCACCGTA 214
 Oy 142 AspPheHisValAlaThrGlyProGly-----GlyGlyAlaGluAlaSer 156
 Db 213 CATTTGGCATGGGGTGTCTTGGCCGAGCGGATGAGACGGGGTGGCGGCTCAACCAAGCT 154
 Oy 157 PheThrAlaProGlyHisThrSerThrPheSerPheLysValLeuGlnProGlyLeuTyr 176
 Db 153 CACATCAAACTGGCGAGACTTCGTCTACAGATTCCAGATGAAGCATAGCGGGAGATT 94
 Oy 177 ValTyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeu 196
 Db 93 ATGTACCAAC---CCGCACTCCGACGAATGTGCAGATGCGCATGATGGCGATG 37
 Oy 197 IleLeuValGluProGlyLeu 203
 Db 36 ATTTGGTGCATTCACCATGAT 16

RESULT	17
CAS554730/c	
LOCUS	513 bp mRNA linear EST 19-NOV-2002
DEFINITION	C8689D07-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus
ACCESSION	CIDNA clone NIA:C0889D07 IMAGE:30033546 5' , mRNA sequence.
VERSION	CAS554730
KEYWORDS	CAS554730.1 GI:25099041
SOURCE	EST.
ORGANISM	Mus musculus (house mouse)
TITLE	Mus musculus
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
COMMENT	Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 513) Pico,Y., Kargull,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Luo,A. and Ko,M.S.H. Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long) Unpublished Contract: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@lysun.gsc.nia.nih.gov Plate: C0889 row: D column: 07 Seq primer: M13 Reverse High quality sequence stop: 513 POLYA=No.

FEATURES	Location/Qualifiers
source	1. .513

```

/organism="Mus musculus"
/mol_type="mRNA"
/etcd_atn="C57BL/6J"
/db_xref="taxon:10090"
/clone="NIA:CO889D07 IMAGE:3003346"
/risau_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_id="NIA Mouse Blastocyst cDNA Library (Long)"
/note="Vector: pSPORT1 (Invitrogen); site_1: SalI; site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://bisun.gic.nia.nih.gov/cDNA). This is

```

a long-transcript-enriched cDNA library (Ref. Genome Res 11: 1553-1558 (2001) [PMID: 11544199]). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo(dT) primer (Invitrogen): 5'-pgACTAGTCTCAGATCGAGCGCGCCCTTTTTTTTTTTT-3') from 0.2 ug of total-RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker lp-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA). "

Alignment Scores:		
Pred. No.:	5.23e-05	513
Score:	145.50	Matches: 45
Percent Similarity:	43.20%	Conservative: 28
Best Local Similarity:	26.63%	Mismatches: 67
Query Match:	5.54%	Indels: 29
Db:	14	Gaps: 8

```

Oy 58 GluIeuProValIleAspAlaIleValIThnHisIaProGluValIProProProValAsp 77
Db 504 GAACGGCGCGATCGACCAAG--CGAGCATGCAGCCACCGTTGATCCACGTCGGGT 448
Oy 78 ArgAspHisProAlaIysValValValI-----LysMetGlu--- 89
Db 447 CCCGACTATGCGCGCGGTGTGACCGCTGAACGGCTGCTGTCGCATGGCGCATGAACGC 388
Oy 90 -----ThValGluIysValMetArg--LeuAlaAspGlyVal 101
Db 387 GACTGAAGAGAGTTTCACTGTCGGCGCAACTGTGGCGCGGATTTGCGGAAGGCGATG 328
Oy 102 GluTyrgInIlePheTrrPheIleGlyGlnValIProGlyGlnMetIleArgValArgIu 121
Db 327 AAGCGTCACTCTGTGGGTTCACMAAGCGGCAATCCCGGGTCCACATTGAGGCACTGA 268
Oy 122 GluAspThrIleGluValGlnPheSerAsnHisProAspSerLysMetProHisAsn--- 140
Db 267 GCGCGAAGCTCCCGATCTTCGTCAACCAAC-----AAATTCGCGGAGCATATCC 220
Oy 141 ---ValAspPheHisAlaAlaIleThrGlyProGly-----GlyGlyAlaGlu 154
Db 219 ACCGTACTACTGGCATGGGTCCTTTTGGCGAGCGGTATGACGGGGTGGGGGCTCAAC 160
Oy 155 AlaSerPheThrAlaProGlyIleHisSerThrPheSerPheLysAlaLeuGlnProGly 174
Db 159 CAGCGTCAATCAACAGCGGAGACCTTGTCTTACGATGCTCCAGATGAACATAGCCGG 100
Oy 175 LeuTyrgValIleHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyr 194
Db 99 ACGTTTATGTATCCAC---CCGCACTCCGACCAAAATGTGCAGATGGCGATGGCATGTG 43
Oy 195 GlyLeuIleLeuValGluProIysGlu 203
Db 42 GGCATGATTTGGTGCATCCAGGTAT 16

```

RESULT 18	CA551823/c	LOCUS	CA551823	515 bp	RNA	linear	EST 19-NOV-2000
DEFINITION	C0849A06-5N NIA Mouse B1a2c0cyc2 CDNA Library (Long)						Mus musculus
	CDNA clone NIA:C0849A06 IMAGE:30029669						5', mRNA sequence.

ORIGIN

Alignment Scores:

Pred. No.: 0.000141 Length: 623
 Score: 142.50 Matches: 47
 Percent Similarity: 44.63% Conservative: 32
 Best Local Similarity: 26.55% Mismatches: 59
 Query Match: 5.43% Indels: 39
 DB: 9 Gaps: 10

US-10-088-045-2 (1-502) x AV822649 (1-623)

```

QY 101 VALGUTGYNPhetp-----ThpneGlygly 110
    |||||.....|
DB 108 GTGGAGTACAGATTGTCGCCGACACAGAGAGCGCGCTTATGACCGTCAAGCGC 167
    |||||.....|
QY 111 GlnvalProGlyGlnmetlleatgvalaGluGlyAspThrIleGluValGlnPheSer 130
    |||||.....|
DB 168 GAGTTCCTGTCGCCACCATTAAGCCTTCCCGAGACACCATCTCGTCAATCTCACC 227
    |||||.....|
QY 131 AennHsProAspSerLysMetProHisAsn-----ValAspPheHisala----- 145
    |||||.....|
DB 228 AAC-----AACTCACACCGAGAGCGCTTGTATCATTCATTCGATGGAATCCGT 275
    |||||.....|
QY 146 -----AlathrGlyProGlyGlyValaGluValaSerPheThrAla 159
    |||||.....|
DB 276 CAGTTCGAGTCCATGCGGAGATGAGACAGACGAGTTCATCAATGCCAATTAAC--- 332
    |||||.....|
QY 160 ProGlyHisThrSerThrPheSerPheLysalaLeuGlnProGlyLeuValTyros 179
    |||||.....|
DB 333 CCGGAGAGACTTTTACCTACATTCATTCGTTGAAAGCGGAGACACATTCCTACCAT 392
    |||||.....|
QY 180 CyalaValaLalaProValaGlyMetHisIleAlaAsnGlyMetTyrosIleLeuVal 199
    |||||.....|
DB 393 -----GSACTATGATGCGATGAGATGAGATGCGCTATTCGATTCGATGATG 443
    |||||.....|
QY 200 GluProLysGlnGlyLeuPro-----LysValAspLysGlyValValMetGln 216
    |||||.....|
DB 444 GACGTGCGCTTAAGAGAAAGACGAGAGATGAGATGATGATGATTAATCTCTTACTC 503
    |||||.....|
QY 217 GlyAspPheTyrosIleLysGly---LysTyrosGlyGlnGlyLeuGlnProPheAspMet 235
    |||||.....|
DB 504 AGTGAATGCTGGATGATGATGATTCCTCCCAAGAACTGCTCTTCT----- 551
    |||||.....|
QY 236 GluLysAlaIleArg-----GluAspAlaGluTyrosValaValaPheAsnGly 250
    |||||.....|
DB 552 TCCAAACTATGCGCTGATCGGTGAAGCTCAGAGCATATTGATTAATGAG 602
    |||||.....|

```

RESULT 26
 CA483187 691 bp mRNA linear EST 14-NOV-2002
 LOCUS LUP12006H1LR LUP12 Linum usitatissimum cDNA clone LUP12006H1LR,
 DEFINITION mRNA sequence.

ACCESSION CA483187
 VERSION CA483187
 KEYWORDS EST.
 SOURCE Linum usitatissimum (Flax)
 ORGANISM Linum usitatissimum

REFERENCE
 AUTHORS Cloutier,S.
 TITLE Flax genomics: ESTs from developing bolls
 JOURNAL Unpublished
 COMMENT Contact: Dr. Sylvie Cloutier
 Cereal Research Centre, Agriculture and Agri-Food Canada
 195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M5
 Tel: (204) 983-2340
 Fax: (204) 983-4604
 Email: scloutier@em.agr.ca

Noti and Mut sites. Not all sequences generated with reverse

primer were from the 5' end (same with forward primer and 3' end).
 Average insert size is 2.07 kb
 Plate: 006 row: H column: 11
 Seq primer: M13 Reverse

FEATURES

source 1..691
 Location/Qualifiers
 /organism="Linum usitatissimum"
 /mol_type="mRNA"
 /cultiivar="Mc Mduff"
 /db_xref="taxon:4006"
 /clone="LUP12006H1R"
 /tissue_type="Bolls"
 /dev_stage="12 days after flowering"
 /lab_host="EMDH10B-70NA"
 /clone_id="LUP12"
 /note="Vector: pCMVSPORTe.1 (Invitrogen Corp.); Site_1:
 Noti; Site_2: Mut; mRNA obtained from bolls 12 days after
 flowering"
 BASE COUNT 170 a 174 c 195 g 152 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.000162 Length: 691
 Score: 142.50 Matches: 49
 Percent Similarity: 41.85% Conservative: 28
 Best Local Similarity: 26.63% Mismatches: 76
 Query Match: 5.43% Indels: 31
 DB: 14 Gaps: 8

US-10-088-045-2 (1-502) x CA483187 (1-691)

```

QY 78 ArgAspHisProAlaLysValValaValaLysMetGluThrValaGlu-----Lys 93
    |||||.....|
DB 108 CCGGTTCACTGTGTATGATGATGATGTCGTGCGCGCCGTCCAGGTTCTGGGGGAAA 167
    |||||.....|
QY 94 ValMetArgLeuAlaAspLysValaGlyValaGlnPheThr----- 107
    |||||.....|
DB 168 ACTAGGCAATTCMAATGGAGCGTGGAGTACATGTTCTGTGTCACCGGACGAGAGAGCAT 227
    |||||.....|
QY 108 -----PheGlyGlyGlnValProGlyGlnMetIleArgValaArgGluGlyAsp 123
    |||||.....|
DB 228 ACTGTCATGCGGTATCAACCGCCAGTTTCCGGGTCCGACGATTAGGCCCAAGTCCGCGAC 287
    |||||.....|
QY 124 ThrIleGluValaGlnPheSerAsn-----HisProAspSerLysMetProHisAsnVal 141
    |||||.....|
DB 288 ATATATCAGTGGACCTCAACCAAGCTCCATACGAGAGGAGCTTATTCAGTGGCAC 347
    |||||.....|
QY 142 AspPheHisAlaAlaThrGlyProGlyGlyValaGluValaSerPheThr----- 158
    |||||.....|
DB 348 GGCATCAGACAGAGAGGAGCTCATGCGGAGATGGGACAGCTGCAATCTCACAGTGCACC 407
    |||||.....|
QY 159 ---AlaProGlyHisThrSerThrPheSerPheLysalaLeuGlnProGlyLeuTyros 177
    |||||.....|
DB 408 ATGGCTCTCGTGGTGAACCTTCCCTTACAGTTCATCTTGACACAGGAGGACATATTTC 467
    |||||.....|
QY 178 TyrHisCysAlaValaLalaProValaGlyMetHisIleAlaAsnGlyMetTyrosIle 197
    |||||.....|
DB 468 TACCAC-----GCCATTACGGGATGCAAGAGTCAAGGAGGCGCTATACGTTTCAATG 518
    |||||.....|
QY 198 LeuValaGluProLysGlnGlyLeuProLys-----ValAspLysGluTyros 212
    |||||.....|
DB 519 ATAGTGAAGCCGGAATATAGTCGGGAGAAAGATCCATTCATTAAGATGCGAGATTC 578
    |||||.....|
QY 213 TyrValaMetGlnLysPheTyrosIleLysGlyValaGluGln---GlyLeuGln 231
    |||||.....|
DB 579 AACCTTGTCTGACGACTGTGGCATTAAGTGTCCAGACCAAAAGTGGGCTCTCT 638
    |||||.....|
QY 232 -----ProPhe 233
    |||||.....|
DB 639 TCCCTCCCTTT 650
    |||||.....|

```

RESULT 27
 AY079863/c


```

QY 53 GINGLHIEGLNGLYGLUeUPROVALIIEAPALIEVALTHRHISALAPROGLUVAL 72
DB 171 -----GATGCCAAGGTGCACAC----- 188
QY 73 PROPROVALIAPARGAPRHISPROVALIYVALIYVALIYMEGLUITHYVALIGU 92
DB 189 -----CACACCTGGGACATCGGTACACATCAACAGTCTCTGCAC 227
QY 93 LYSVALMECARGLEUVALAPRGYVALIGUTYRGINPHETRPHETGLYGINVAL 112
DB 228 TCCGCTCAACAGCTGCAG-----GTGACCATCAACCGCGAGTCC 266
QY 113 PROGLYGLMEGLIYRGYVALARGGLUGIYAPRTHIEGLIYVALINPHESERAHNIS 132
DB 267 CCGGGCCGACCATTCGCGCCGACGCGGACAGCTGGTGTACCGCTGCCAACAAGT 326
QY 133 PROAPSERLYMECPROHISAPENVALAPRPHENIHALALATHRGILYPROGLYGLY 151
DB 327 CTGGACACAGAAACAC-----GGCATTCACCTGSCACGGATCCGCGAGTCCGCGG 383
QY 152 -----GTYALAGIUALASERPHETHALAPROGLIYHISRHSETHR 165
DB 384 TGGGCGGACGCGACCGTCCGCGCTCACGAGTGTCCATCCTCCCGCGGACACTTCACC 443
QY 166 PHESERPHELYALALEUGINPROGLYGLUETHYVALIYTHIEGYALVALIAPROVAL 185
DB 444 TACAGGTGTCTGTCCACCGGCGACCTACTTCTTACAC-----GCCCATCTAT 494
QY 186 GLYMERHIEIIEALAEANGLYMECTYRGYLYLEUVALIGLUPROLYSGIUGLYLEU 205
DB 495 GGGATTCACCGCGTGGCGCGGCTCGACGGATCTGTGTGTGTGTGTGTGTGTGTGTGT 554
QY 206 PROLYS-----VALAPLYSGUTYTYTYVALMEGLINGLYAPRPHETRYHLYS 222
DB 555 GCGGACCGCTTCCCTACGACGACGACGACGACGACGACGACGACGACGACGACGACG 614
QY 223 GLYLYTYRGYGLUINGLYLEUGINPROPHAPRMEGLIYVALIIEARGIUALAP 242
DB 615 AGCGTGTAC-----GAGCAGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 644
QY 243 ALAGLUTYVALIYALAPHEANGLYSERVALIYALALEU 255
DB 645 TCCGACCGCTCTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 683

```

RESULT 32
 CASS3237 515 bp mRNA linear EST 19-NOV-2002
 LOCUS C0869G05-5N NIA Mouse Blastocyst cDNA Library (Long) Mus musculus
 DEFINITION CDNA clone NIA:C0869G05 IMAGE:30031660 5', mRNA sequence.

ACCESSION CASS3237
 VERSION CASS3237.1 GI:25097478
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 515)
 Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Tanaka, T., Luo, A. and
 Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)
 JOURNAL Unpublished
 COMMENT Contact: Dawood B. Dudekula

Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun-grc.nia.nih.gov
 Plate: C0869 Row: G column: 05
 Seq primer: M13 Reverse
 High quality sequence stop: 515
 POLYANo.

FEATURES
 source Location/Qualifiers
 1..515

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="nlaEST:C0869G05-5N"
/db_xref="taxon:10090"
/clone="NIA:C0869G05 IMAGE:30031660"
/tissue_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_id="NIA Mouse Blastocyst cDNA Library (Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun-grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). PMID: 11544199). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo (dT) primer (Invitrogen): 5'-DGACTAGTTCATGATCGGAGCGCGCCCTTTTCTTTT-3' from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol precipitation. The cDNAs were ligated to lone-linker L1-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

```

BASE COUNT 100 a 156 c 151 g 108 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.000176 Length: 515
 Score: 140.50 Matches: 45
 Percent Similarity: 42.60% Conservative: 27
 Best Local Similarity: 26.63% Mismatches: 68
 Query Match: 5.35% Indels: 29
 DB: 14 Gaps: 8

US-10-088-045-2 (1-502) x CASS3237 (1-515)

```

QY 58 GULUeUPROVALIIEAPALIEVALTHRHISALAPROGLUVALPROPROVALAP 77
DB 504 GAAGCGCGCATCATGACAAAG---GCGACGATGACGCCACCGTGCATCCAGTCCGCT 448
QY 78 ARGAPRHISPROVALIYVALIYVALI-----LYMEGLU----- 89
DB 447 CCGGACTATCGCCCGGTGTGTGACGCTGAACGCGTGTGTGTGTGTGTGTGTGTGTGT 388
QY 90 -----THYVALIGIYVALMECARG---LEUVALAPGIYVAL 101
DB 387 GAGTGAAGAGATTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 328
QY 102 GLUTYRGINPHETRPHETGLYGLIYVALIYVALIYGLIYGLIYGLIYGLIYGLIY 121
DB 327 AAGGCTCATCTGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 268
QY 122 GLYAPRTHIEGLIYVALINPHESERAHNISPROAPSERLYMECPROHISAPEN 140
DB 267 GGGGACAGCTCGTATCTTCTGTACCAAC-----AAATTGCCGAGCATTACC 220
QY 141 ---VALASPHENIHALALATHRGILYPROGLY-----GLYIYALAGIU 154
DB 219 ACCGTATATGGATGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 160
QY 155 ALASERPHETHALAPROGLIYHISRHSETHRPHESERPHELYALALEUGINPROGLY 174
DB 159 CAGCTCATCAACCTGCGGACACTTGTCTACGAGTTCAGATGGAAGCATACGCGG 100

```


FEATURES	source
1. .614	location/Qualifiers
/organism="Arabidopsis thaliana"	
/mol_type="mRNA"	
/db_xref="taxon:3702"	
/clone="RAFL08-11-P16"	
/dev_stage="rosette plants"	
/lab_host="DH10B"	
/clone_idb="RAFL08"	
/note="Site 1: BamHI, Site 2: SalI; subjected to dehydration-treated (1, 2, 5, 10, 24 hr)"	
161 a 140 c 166 g 146 t 1 others	
ORIGIN	
BASE COUNT	
161 a 140 c 166 g 146 t 1 others	
ALIGNMENT SCORES:	
Pred. No.:	0.000265
Score:	139.50
Percent Similarity:	44.07%
Best Local Similarity:	26.55%
Query Match:	5.31%
DB:	9
Gaps:	10
US-10-088-045-2 (1-502) x AV826269 (1-614)	
Y	101 ValGluTyrGlnPheTrp-----ThrPheGlyGly 110
Db	105 GTGAGAGTACAGTATGTCGCCGCGAGCTGCAAAAGGCGCCGTATATGACCGTCAAGCGC 164
Y	111 GlnValProGlyAlaMetLeuArgValArgGluGlyAspThrIleGluValGlnPheSer 130
Db	99 ACGTTTAGTACAC---CCGACTCCGACGAATATGTCAGATGCGATGGCATGGATG 43
Y	195 GlyLeuIleLeuValGluProGlyGlu 203
Db	42 GGCATGATGTGCTGCATCCACGATAT 16
RESULT 33	
AV826269	614 bp mRNA linear EST 01-APR-2002
LOCUS	AV826269 RAF08 Arabidopsis thaliana cDNA clone RAFL08-11-P16 5',
DEFINITION	mRNA sequence.
ACCESSION	AV826269
VERSION	AV826269.1 GI:19868329
KEYWORDS	EST.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucoside 11; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS	1 (bases 1 to 614)
TITLE	Seki, Y., Narasaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
COMMENT	large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL	Unpublished
CONTACT	Contact: Motoaki Seki
PLANT FUNCTIONAL GENOMICS RESEARCH GROUP	
RIKEN GENOMIC SCIENCES CENTER	
3-1-1 KOYADAI, TSUKUBA, IBARAKI 305-0074, JAPAN	
TEL: 81-298-36-4359	
FAX: 81-298-36-9060	
EMAIL: msek@rc.riken.go.jp	
AN ARABIDOPSIS FULL-LENGTH CDNA LIBRARY WAS CONSTRUCTED ESSENTIALLY AS REPORTED PREVIOUSLY (SEKI ET AL., 1998). CDNA CLEAVED WITH BAMHI AND XHOI WAS LIGATED TO MODIFIED LAMBDA P/LC-1 VECTOR (CARNINCI ET AL., SUBMITTED FOR PUBLICATION) DIGESTED WITH BAMHI AND SALI. THIS CLONE IS IN A MODIFIED PHAGESCRIPT VECTOR. PLEASE VISIT OUR WEB SITE (HTTP://WWW.GSC.RIKEN.GO.JP/E/PLANT/INDEX_E.HTML) FOR FURTHER DETAILS.	

DB	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Db	165	GAGTTCCTGGTCCACCATTAAGACCTTGGCCGAGACACATCGTCGCAATTCACC	131	AsnH1bProApspSerlyMetProH1aAsn-----ValaSpH1a-----	145
Db	225	AAC-----AACTCACACCGAAGCCCTTGCTCACTCATGGCAGATTCGCT	146	AlathGlyProGlyGlyAlaGlyAlaSerPheThraLe	159
Db	273	CAGTTCCGAAGTCCATGGCAGATGGAGCAGACAGAGTACTCATATGGCATTAAAC--	160	ProGlyH1aThSerThrPheSerPheLysAlaLeuGlnProGlyLysLeuValTyrHis	179
Db	330	CCTGGAGAGACTTTACTCTACATTTACATTTACATTTACATTTACATTTACATTTACAT	180	CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLysIleLeuVal	199
Db	390	-----GGACACTATGGCATGCACAGATCAGACTGCTATACGATCGCTTATGTCG	200	GluProLysGlyGlyLysLeuPro-----LysAlaAspLysGlyTyrTyrValMetGln	216
Db	441	GACCTGGCTTAAGAAAGAGAGCAGAGATTGAGATACATGGTGTGATTATCTCTTACTC	217	GlyAspPheTyrThrLysGlyLys-----LysTyrGlyGlyGlnGlyLeuGlnProPheAspMet	235
Db	501	AGTACACTGCTGGCATGAGCTATTCCTCCACAGAACTCGTCTTTCT-----	236	GlyLysAlaIleArg-----GluAspAlaGlyTyrValValPheAsnGly	250
Db	549	TCCAAACTTATGCCCTGATCGGTGAGCTCAGAGCTCAGATGATTAATGATAAATGGG	549	TCCAAACTTATGCCCTGATCGGTGAGCTCAGAGCTCAGATGATTAATGATAAATGGG	599
RESULT 34	LOCUS	EG48142	659 bp	mRNA	linear
DEFINITION	Accession	BF06088C1F1066	Elicited cell culture	Medicago truncatula	CDNA
KEYWORDS	Version	BF06088C 5'	mRNA sequence.		
ORGANISM	Source	EG48142.1	GI:13366923		
		EST.			
		Medicago truncatula (barrel medic)			
		Medicago truncatula			
		Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
		Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; rosids			
		; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;			
		Medicago.			
		1 (bases 1 to 659)			
		Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,			
		Flores,H.R., Iman,J.T., Weller,J.W. and May,G.D.			
		Expressed Sequence Tags from the Samuel Roberts Noble Foundation -			
		Center for Medicago Genomics Research			
		Unpublished			
		Contact: Dixon RA			
		Plant Biology Division			
		The Samuel Roberts Noble Foundation			
		2310 Sam Noble Parkway, Ardmore, OK 73402, USA			
		Tel: 580 221 7302			
		Fax: 580 221 7380			
		Email: radixon@noble.org			
		Insert Length: 659 Std Error: 0.00			
		Plate: 106 row: G column: 08			
		Seq primer: TCACACAGGAAACAGCTATAC.			
FEATURES		Location/Qualifiers			
Source		1..659			
		/organism="Medicago truncatula"			
		/mol_type="mRNA"			
		/db_xref="taxon:3880"			
		/cclone="NF106G08EC"			
		/tissue_type="Cell cultures derived from root tissues"			
		/dev_stage="Cell suspensions were subcultured every 14			
		days. Cells were induced six days after subculture"			
		/clone_lib="Elicited cell culture"			
		/note="Vector: Lambda Zap; Cells were induced with yeast			
		cell wall extracts equivalent to 50ug/ml glucose in the			
		final concentration. Samples were taken at 0.5, 1, 12 and			

24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 0.000314	659	59	29	88	50	10
Percent Similarity: 139.50						
Best Local Similarity: 38.94%						
Query Match: 26.11%						

US-10-088-045-2 (1-502) x BG448142 (1-659)

```

QY 24 SerAnGlnAlaAspLysAlaGlnProLysSerSerThrValAspAlaAlaLys 43
DB 86 TCTAGACATGCTCACTCACTCAAGGCGACCAAAATTATTGTTCTATG 145
QY 44 ThrAlaAsnAlaAspAlaAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGln 63
DB 146 TTTCTTTCT-----CATTCGTGTAATTTCCAAAGCTGA 181
QY 63 palAlaIleValThrHisAlaProGluValProProValAspArgAspHisProAlaLys 83
DB 182 GCGTGGCTGTTAGACAT----- 197
QY 83 sValValValLysMetGlnThrValGlnLysValMetArgLysAlaAspGlyValGlnLys 103
DB 198 -----TATTAATGGAG--GTGAAGTATGATTATAGATCCCTGATGTTATAGAA 247
QY 103 rGlnPheTrpThrPheGlyGlnValProGlyGlnMetIleArgValArgGlnGlnLys 123
DB 248 ACTTGTTATACATTAAATGAAGAGCTCAGAGCTCACTCAAGCAGACAGAGGCTGA 307
QY 123 pThrIleGlnValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspHis 143
DB 308 TACTGTTGATGTTCAACTTAACAAC-----AAATGCTCAGAAAACCTGCCAT 358
QY 143 eHis-----AlaAlaThrGlyProGlyGlnGlnLys 153
DB 359 CCATGGCAGTGTATTAGACAGATTGGAACTCTGTTGATGAGAACAGAGAGTGTCT 418
QY 153 aGluAlaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaGlnLys 173
DB 419 TCATATGCTTATA---CTACCTGGAGACACCTTTGTTATGATTTGTTGATAGGCC 475
QY 173 oGlyLeuTrpValThrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMet 193
DB 476 TGGTACATATCTTATCAT-----GCTCAGTATGAGAAAGGAGAGAGT 526
QY 193 tTyrGlyLeuIleLeuValGlnProLysGlnGlyLeuPro--LysValAspLysGlnLys 212
DB 527 ATATGAGATGATGCTGTGCGACCTAATGAGCCCTGAACTTTTCTTATGATGATAG 586
QY 212 rTyrValMetGlnGlnLysPheTrpThrLysGlyLysTrpGlyGlnGlnGlnGlnGln 231
DB 587 AAGCATTAATTTGATGATGTTGATGATGATGATGATGATGATGATGATGATGAT 643
QY 232 -----ProPhe 233
DB 644 GTCTGCAATCCCTTTC 659

```

RESULT 35

BF644576 690 bp mRNA linear EST 20-DEC-2000
 LOCUS NF015H09EC1079 Elicited cell culture Medicago truncatula cDNA
 DEFINITION clone NF015H09EC 5', mRNA sequence.
 ACCESSION BF644576
 VERSION BF644576.1 GI:11909705
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosoid I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.

REFERENCE

1 (bases 1 to 690)
 Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
 Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.
 Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
 Center for Medicago Genomics Research

JOURNAL

Unpublished
 Contact: Dixon RA
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7302
 Fax: 580 221 7380
 Email: radixon@noble.org
 Insert length: 690 Std Error: 0.00
 Plate: 015 row: H column: 09
 Seq primer: TCACAGAGAAACGCTATGAC.
 Location/Qualifiers

FEATURES

source
 1..690
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /db_xref="taxon:3880"
 /clone="NF015H09EC"
 /tissue_type="Cell cultures derived from root tissues"
 /dev_stage="Cell suspensions were subcultured every 14
 days. Cells were induced six days after subculture"
 /note="Vector: lambda Zap; Cells were induced with yeast
 cell wall extracts equivalent to 50ug/ml glucose in the
 final concentration. Samples were taken at 0.5, 1, 12 and
 24 hours after induction. Equal amounts of RNA from each
 time point were pooled and used for mRNA isolation."

BASE COUNT 202 a 136 c 135 g 217 t

ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 0.000335	690	59	29	88	50	10
Percent Similarity: 139.50						
Best Local Similarity: 38.94%						
Query Match: 26.11%						

US-10-088-045-2 (1-502) x BF644576 (1-690)

```

QY 24 SerAnGlnAlaAspLysAlaGlnProLysSerSerThrValAspAlaAlaLys 43
DB 87 TCTAGACATGCTCACTCACTCAAGGCGACCAAAATTATTGTTCTATG 146
QY 44 ThrAlaAsnAlaAspAlaAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGln 63
DB 147 TTTCTTTCT-----CATTCGTGTAATTTCCAAAGCTGA 182
QY 63 palAlaIleValThrHisAlaProGluValProProValAspArgAspHisProAlaLys 83
DB 183 GCGTGGCTGTTAGACAT----- 198
QY 83 sValValValLysMetGlnThrValGlnLysValMetArgLysAlaAspGlyValGlnLys 103
DB 199 -----TATTAATGGAG--GTGAAGTATGATTATAGATCCCTGATGTTATAGAA 248
QY 103 rGlnPheTrpThrPheGlyGlnValProGlyGlnMetIleArgValArgGlnGlnLys 123
DB 249 ACTTGTTATACATTAAATGAAGAGCTCAGAGCTTACCATTCACAGCAGAGGCTGA 308
QY 123 pThrIleGlnValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspHis 143
DB 309 TACTGTTGATGTTGAAGTTAAACAAC-----AAATGCTCAGAAAACCTGCCAT 359
QY 143 eHis-----AlaAlaThrGlyProGlyGlnGlnLys 153

```

Db 360 CCATTGGCATGTATTAGACAGATTGGAACCTCTGTTTGTATGAAACAGAAAGAGTGC 419
 153 aGuaLaSerPheThrAlaProGlyHisThrSerThrPheSerPheLysAlaLeuGlnPr 173
 Db 420 TCACATCTCTATA---CTACCTGGAGACACCTTTGTTTATCGATTGTTGTTGATAGGCC 476
 173 oGlyLeuTyValTyHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMe 193
 Db 477 TGTACATATCTTTATCAT-----GCTCAGTATGAAATGCAAGGAGACAGAGT 527
 193 tTyrgGlyLeuIleLeuValGluProLysGlyLeuPro---LysValAspLysGly 212
 Db 528 ATATGGAATGATTGCTGGACACCTATAGACCTGAACCTTTTCTTATGACTTGTATAG 587
 212 tTyValMetGlnGlyAspPheTyThrLysGlyLysTyrgLysGluGlnGlyLeuGln-- 231
 Db 588 AAGCATATTTGATGATGATGGTACCATAGAGTACTTAT--GAACAACTGCTAGATT 644
 232 -----ProPhe 233
 Db 645 GTCTGCATCCCTTTT 660

RESULT 36
 AV826844 571 bp mRNA linear EST 01-APR-2002
 LOCUS AV826844 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-07-M21 5',
 mRNA sequence.
 ACCESSION AV826844
 VERSION AV826844.1 GI:19868904
 SOURCE EST.
 ORGANISM Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 571)
 Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Saitou,M., Nakajima,M.,
 Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
 Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
 and Shinozaki,K.
 Large scale analysis of Arabidopsis full-length cDNA (2002b)
 TITLE
 JOURNAL Unpublished
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: mseki@tc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
 and XhoI was ligated to modified lambda FLC-1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified Bluescript vector. Please visit our web
 site (http://www.gsc.riken.go.jp/e/planet/index_e.html) for further
 details.

FEATURES
 source
 1..571
 /location/Qualifiers
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL09-07-M21"
 /dev_stage="plants at various developmental stages from
 germination to mature seeds"
 /lab_host="DH10B"
 /clone_1ib="RAFL9"
 /note="Site 1: BamHI; Site 2: SalI; subjected to
 dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
 hr) treatments"
 145 a 133 c 154 g 136 t 3 others

BASE COUNT 145 a 133 c 154 g 136 t 3 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.000291 Length: 571
 Score: 139.00 Matches: 39
 Percent Similarity: 45.14% Conservative: 26
 Best Local Similarity: 27.08% Mismatches: 47
 Query Match: 5.29% Indels: 32
 DB: 9 Gaps: 7

US-10-088-045-2 (1-502) x AV826844 (1-571)
 101 ValGluTyrglnPheTrp-----ThrPheGlyGly 110
 105 GTGAGGTACAGATATTGCTGCCGAGCTGCAAGAGCGCGCTTATAGCCTCAACGGC 164
 111 GlnValProGlyGlnMetIleArgValArgGluGlyAspThrIleGluValGlnPheSer 130
 165 GAGTTTCTGTGTCACCATTAAGCCTTCCGCGAGACACCATGCTGCTCAATCTCAC 224
 131 AenHisProAspSerLysMetProHisAsn-----ValAspPheHisAla----- 145
 225 AAC-----AACTCACACCGAAGCCTTGTCATCATTCATTTGGCATGGAATCCGT 272
 146 -----AlaThrGlyProGlyGlyGlyAlaGluAlaSerPheThrAla 159
 273 CAGTTCGGAAGTCCATGGGACAGATGAGCAGCAGAGTACTCAATGCGCAATTAC-- 329
 160 ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLeuTyValTyHis 179
 330 CCGAGAGACACTTATTAACATTTCACTGTTGAAAGCCGGGAACACATTTCTACAT 389
 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrgLysLeuVal 199
 390 -----GACACTATAGGATGAGATGAGATGAGTGGGCTTATAGCATCGTTATGG 440
 200 GluProLysGluGlyLeuPro-----LysValAspLysGluTyValTyValMetGln 216
 441 GAGCTGGCTAAAGAAAGACGAGATGATGATGATGATGATGATGATGATGATGATG 500
 217 GlyAspPheTyR 220
 501 ACTGACTGCTGG 512

RESULT 37
 CA553614 489 bp mRNA linear EST 19-NOV-2002
 LOCUS CA553614/C
 DEFINITION C0874E11-5N NIA Mouse Blastocyst cDNA library (long) Mus musculus
 C0874E11 IMAGE:30032122 5', mRNA sequence.
 ACCESSION CA553614
 VERSION CA553614.1 GI:25097868
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 489)
 Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Luo,A. and
 Ko,M.S.H.
 TITLE
 JOURNAL Systematic Analyses of NIA Mouse Blastocyst cDNA Library (Long)
 COMMENT Unpublished
 CONTACT Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: C0874 row: B column: 11
 Seq primer: M13 Reverse
 High quality sequence stop: 489
 POLYA-No.

FEATURES
 source
 1..489
 /location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"

```

BASE COUNT
ORIGIN

99 a      144 C      144 G      102 T

/db_xref="niaEST:C0874E11-5N"
/db_xref="taxon:10090"
/clone="NTA:C0874E11 IMAGE:30032122"
/clone_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_idb="NTA Mouse Blastocyst cDNA Library (Long)"
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NTA), Intramural Research
Program, NIH (http://lgsun.grc.nia.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001)). [PMID: 11544199]". Total RNAs were
extracted from a pool of 20 Blastocysts. Double-stranded
cDNAs were synthesized with an Oligo(dT) primer.
[Invitrogen:
5'-GGACTAGTCTAGATCCGAGCGCGCCGCTTTTTTTTTTTT-3'] from
0.2 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-5. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.2 kb. The library was constructed
by Yulan Piao (NTA)."

```

Alignment Scores:					
Pred. No.:	0.000266	Length:	489		
Score:	138.50	Matches:	35		
Percent Similarity:	47.50%	Conservative:	22		
Best Local Similarity:	29.17%	Mismatches:	50		
Query Match:	5.27%	Indels:	13		
DB:	14	Gaps:	5		

US-10-088-045-2 (1-502) x CA553614 (1-489)	
Oy	92 GluIysValMetArg---LeuAlaSerPheValGluIyrGlnPhetRhpPheGlyLy 110
Db	360 GAGCCGGCTGTCCCGCAATTTCGCCAGSGAATGAAGCTCACAATTATGGGGCATTAACGCT 301
Oy	111 GlnValProGlyGlnMetLleArgValArgGluGlyAspThrLleGlnValGlnPheSer 130
Db	300 CAATGC CGGGCCCCACCATTGAGCCGCTCGAAAGCGCAGAAGCTCCGACTTTCGTCACG 241
Oy	131 AsnHisProAspSerLysMetProHis-----AsnValAspPheHisAlaAlaThrCly 148
Db	240 AAC-----AAACTGCCGAMACAGACGACCATTCATTGGCACAGTGTCCTCTG 193
Oy	149 ProGly-----GlyGlyAlaGluAlaSerPheThrAlaProGlyHisThr 163
Db	192 CCCAACGGCATGAGACGGCGTTGTGTGACTCAATCAGCCCCACATTCCCCCGGACAGACC 133
Oy	164 SerThrPheSerPheLysAlaLeuGlnProGlyLeuTyrrValTYrHisCyseAlaValAla 183
Db	132 TTCGTGTACGACTTCAGATGACAGACACGGGAGCTTCATGTATCAT---CCGATTCTG 76
Oy	184 ProValGlyMetHisLleAlaAsnGlyMetYrrGlyLleLleValGlnProLysGln 203
Db	75 GACGAATAATGCTCCAGATGGCCGATGGCAGTATGAGGATGATGCTGTCATCCGGCGCAT 16

RESULT 38	
BF650165	645 bp mRNA linear EST 20-DEC-2000
LOCUS	NF088G08EC1F1066 Elicited cell culture Medicago truncatula cDNA
DEFINITION	clone NF088G08DEC 5', mRNA sequence.
ACCESSION	BF650165

```

VERSION      BF650165.1  GI:11915295
KEYWORDS     EST.
SOURCE       Medicago truncatula (barrel medic)
ORGANISM     Medicago truncatula
              Birkarvete; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
              eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
              Medicago.
REFERENCE     1 (bases 1 to 645)
AUTHORS      Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
              Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.
TITLE        Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
              Center for Medicago Genomics Research
JOURNAL      Unpublished
COMMENT      Contact: Dixon RA
              Plant Biology Division
              The Samuel Roberts Noble Foundation
              2110 Sam Noble Parkway, Ardmore, OK 73402, USA
              Tel: 580 221 7302
              Fax: 580 221 7380
              Email: radixon@noble.org
              Insert Length: 645 Std Error: 0.00
              Plate: 088 row: G column: 08
              Seq primer: TCACACGAAGAACGACATGAC.
FEATURES     Location/Qualifiers
              1..645
SOURCE

```

```

/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
days. Cells were induced six days after subculture"
/clone_id="Elicited cell culture"
/notes="Vector: Lambda Zap; Cells were induced with yeast
cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."
BASE COUNT      189 a      116 c      136 g      203 t      1 others
ORIGIN
Alignment Scores:
  Pred. No.:      0.000438      Length:      645
  Score:      138.00      Matches:      45
  Percent Similarity:      40.96%      Conservative:      23
  Best Local Similarity:      27.11%      Mismatches:      58
  Query Match:      5.26%      Indels:      40
  DB:      10      Gaps:      8
US-10-088-045-2 (1-502) x BF650165 (1-645)
QY      100 G|yValG|utyrg|nPh|eTp-----106
      |||||      |||
      ::      ::
Db      90 GGTGTAGACATTTAAATGGAGCGTAGATGATTAAGATCCCGATTGTTATAG 149
      |||||      |||||
      ::      ::
QY      107 -----ThrPheG|yG|nValP|roG|yInMeI|eArGValArG|uG|y 122
      |||||      |||||      |||||      |||||
      ::      ::      ::      ::
Db      150 AAACCTGTATTAACCATTAATGGAAGAAGCTCCAGACCTACCATTCAGACAGAGGGGT 209
      |||||      |||||      |||||      |||||
      ::      ::      ::      ::
QY      123 AsPThrI|eG|uValG|nPh|eSerAsnHisPro|aPse|rlyS|meC|ProHisAsnValAsp 142
      |||||      |||||      |||||      |||||
      ::      ::      ::      ::
Db      210 GATCTGTGTGTAGTTGAAGTTAAAC-----AAATGTCCTACAGAAACCTTGCC 260
      |||||      |||||
      ::      ::
QY      143 PheHis-----Ala|a|aTrnG|yP|roG|yG|y 152
      |||||
      ::
Db      261 ATTCATTGGCAGTGTATTAGACAGATTGGACCTCTTGTTGATGGAACAGAGAGACTG 320
      |||||
      ::
QY      153 A|aG|u|a|a|e|rPhe|Tr|a|a|P|roG|y|n|H|e|Th|e|Tr|Phe|Se|rPhe|y|a|a|e|u|G|n 172
      :::::      |||||      |||||      :::::
Db      321 TCTCATGTCTCTATA---CTACCTGGAGACACCTTTGTTTATGCATTTGTTGTCATAGG 377
      :::::      |||||      |||||      :::::
QY      173 ProG|yLeuTyGValTyGn|e|Cy|a|a|Val|a|a|ProVal|a|G|yMe|e|n|I|a|a|a|n|G|y 192

```

```

||||| ||||| ||||| |||||
Db 378 CTTGTAACATCTTATCAT-----GCTCACTATGAATCCAGGAAGCAGGA 428
Oy 193 MetTyrGlyLeuIleLeuValGluProLysGluGlyLeuPro---LysValAspLysGlu 211
Db 429 GATATGGAATGATTCGTGGACCTATGACCCCGAACCTTTTCTTATGACTTTCAT 488
Oy 212 TyrTyrValMetGlnGlyAspPheTyrThrLysGlyLysTyrGlyGluGlnGlyLeuIle 231
Db 489 AGAAGCATATTTTGAATGATTCATGACATAGAGTACTTAT---GAACATCTGCTTGA 545
Oy 232 -----ProPhe 233
Db 546 TTGCTGCAATCCCTTTT 563

RESULT 39
BU635352 737 bp mRNA linear EST 23-SEP-2002
BU635352
LOCUS 020H08 Infected Arabidopsis leaf Arabidopsis thaliana cDNA, mRNA
DEFINITION
sequence.
ACCESSION BU635352.1 GI:23302607
VERSION BU635352
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 737)
Lundgaard, M., Emmersen, J., Nielsen, K.L., Wilson, I., Somerville, S.
and Weidner, K.G.
EST sequencing of Erysiphe cichoracearum infected Arabidopsis
plants
JOURNAL Unpublished
COMMENT Contact: Karen G. Weidner
Institute for bioteknologi
Aalborg Universitet
Sohngaardsholmvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808
Email: kweid@bio.auc.dk
Location/Qualifiers
FEATURES
source
1. 737
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/db_xref="taxon:3702"
/dev_stage="Plant 3 weeks old, three days post infection"
/clone_id="Infected Arabidopsis leaf"
/note="Organ: Leaf; Vector: pBluescript; Mixed cDNA
library of Arabidopsis and E. cichoracearum infected leaf
from three weeks old Arabidopsis plants. Plants were
harvested 3 days after infection and mRNA oligo dt
selected."
BASE COUNT 208 a. 164 c 164 g 201 t
ORIGIN
Alignment Scores: 0.00067 Length: 737
Pred. No.: 137.00 Matches: 67
Score: 37.55% Conservative: 28
Percent Similarity: 37.55% Mismatches: 96
Best Local Similarity: 26.48% Indels: 62
Query Match: 5.22% Gaps: 13
DB: 13

US-10-088-045-2 (1-502) x BU635352 (1-737)
Oy 107 ThrPheGlyGlyLeuValProGlyGlnMetIleArgValArgGluGlyAspThrIleGlu 126
Db 25 ACAATCAATGATGTTCCAGTCCCACTTAAGCTCAACAAAGGTGACACCATCTT 84
Oy 127 ValGlnPheSerAsnHisProAspSerLysMetProHisAsnValAspPheHis----- 144
||||| ||||| ||||| |||||

```

```

Db 85 GTGAGCTCAAGAT-----ACCTTCATGACTGAAAAATGCTGCTCCATGGCAT 135
Oy 145 -----AlaIleThrGlyProGlyGlyValAlaGluAlaSerPheThrAla---- 159
Db 136 GGAATCGACAGATTGGGACTCCATGCTTCATGGAGTGAAGGTCTTACTCAATGTCA 195
Oy 160 -----ProGlyHisThrSerThrPheSerPheLysAlaLeuGlnProGlyLysLeuTyrVal 177
Db 196 ATCTTCCTCGAAGAGCTTCATTATACCAATTGCTGTTGATAGGCTGTCATACATG 255
Oy 178 TyrHisCysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIle 197
Db 256 TATCAT-----TCACACTATGGATGCAGAGAGAGTCTGATTAATCGGAATGATT 306
Oy 198 LeuValGluProLysGluGlyLeuProLys-----ValAspLysGluTyrTyrVal 214
Db 307 CAAGTTTCTCTCCGCGCACAGAGCCGAACCGTTATACATGATTAATGACGAACTTT 366
Oy 215 MetGlnGlyAspPheTyrThrLysGlyLysTyrGlyGlu---GlnGlyLeuGln----- 231
Db 367 TTGTTAAGTAACTGATTCACAAAGATGATGTCAGAGAAAGCCACCGGTTAGCATCATA 426
Oy 232 ProPheAspMetGluLysAlaIleArgGluAspAlaGluTyrValVal----- 247
Db 427 CCCTTC-----AAGTGGTCCGTGAGCCCAATCCGTTATGATACAGAGAGAGA 477
Oy 248 ---PheAsnGlySer-----ValGlyAlaLeuThr 256
Db 478 AGATTCAAGCTCTCAAAACAACCTTACCACTCTCCAGCTTATGTCGTGAGATGTATAC 537
Oy 257 GlyLysLeuAla-----LeuLysAlaLysValGlyGluThrValArg 270
Db 538 GTCTCAACCGTGAATGTTGCGGTTTATTTTACGGTATCCCGGAAAGACATACAG 597
Oy 271 LeuPheValGlyAsnGlyGlyProAsnLeuThrSerSerPheHisValIleGlyLysIle 290
Db 598 CTGCAATAGCTACTGATGAGCGCTCTGCTCTCAGTTTCCAAATC----- 645
Oy 291 PheAspLysValHisPheGlnGlyLysGlyLysLeuAsnHisAsnIleGlnThrLeu 310
Db 646 -----GAGGACACTAATTTG-----ACAGTT 666
Oy 311 IleProAlaGlyGlyAlaIleThrGluPheLysVal 323
Db 667 GTTGAAGCAGAGGACACTTACTGATACCATTTACCGTG 705

RESULT 40
AV826343 612 bp mRNA linear EST 01-APR-2002
AV826343
LOCUS AV826343 RAF18 Arabidopsis thaliana cDNA clone RAFL08-13-A06 5',
DEFINITION
mRNA sequence.
ACCESSION AV826343.1 GI:19868403
VERSION AV826343
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 612)
Seki, M., Narusaka, M., Ishida, Y., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

```

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda FL-C-1 vector (Garnier et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source

1..612
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF08-13-A06"
/dev_stage="rosette plants"
/lab_host="DH10B"
/clone_id="RAF08"
/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration-treated (1, 2, 5, 10, 24 hr)"

BASE COUNT 160 a 140 c 165 g 144 t 3 others

ORIGIN

Alignment Scores:

Prod. No.: 0.000586 Length: 612
Score: 136.50 Matches: 46
Percent Similarity: 44.07% Conservative: 32
Best Local Similarity: 25.99% Mismatches: 60
Query Match: 5.20% Indels: 39
DB: 9 Gaps: 10

US-10-088-045-2 (1-502) x AV826343 (1-612)

QY 101 ValGluTyrGlnPheTTP-----ThrPheGlyGly 110
DB 105 GTGGAGTACAGATGATGTCGCGAGTGCAGAGGCGCGTTATGACCGTCAACGGC 164
QY 111 GluValProGluGlnMetLeuValArgGluGluValPheThrIleGluValGlnPheSer 130
DB 165 GAGTTTCTGCTCCACCATTAAGCCTTGCGGAGACACCATCGTCGTCATCTCACC 224
QY 131 AsnHisProAspSerLeuMetProHisAsn-----ValAspPheHisAla----- 145
DB 225 AAC-----AAATCACCACCGAAGCGCTTGTCTCATCTGCGATGGATCGGT 272
QY 146 -----AlaThrGlyProGlyGlyValAlaGluAlaSerPheThrAla 159
DB 273 CAGTTCCGAGTCCATGGCGAGATGAGACAGAGAGTACTCAATGCCAATTAC-- 329
QY 160 ProGluHisThrSerThrPheSerPheLeuAlaLeuGlnProGluLeuTyrValTyrHis 179
DB 330 CCTGGAGAGACTTTTACTTACATTTTCACTGTAAGAACCGGAGACACATTTTCTACAT 389
QY 180 CysAlaValAlaProValGlyMetHisIleAlaAsnGlyMetTyrGlyLeuIleLeuVal 199
DB 390 -----GGACACTATGGCATGCAGATCAGTGGGCTATAGCGATCGTGTATTGTG 440
QY 200 GluProLysGluGlyLeuPro-----LysValAspLysGluTyrTyrValMetGln 216
DB 441 GACGTGCTTAAGGAAAGAGGAGATGAGATGACATGATGATGATTAATCTTACTC 500
QY 217 GlyAspPheTyrThrLysGly--LysTyrGlyGluGlnGlyLeuGlnProPheAspMet 235
DB 501 AGTACTGTGGCATGAGCTATTCCTCCCAAGAACTCGGCTTCTC----- 548
QY 236 GluLysAlaIleArg-----GluAspAlaGluTyrValValPheAsnGly 250
DB 549 TCCAAACTATGCGCTGATCGTGAAGCTCAGACATATTGATTAATGGG 599

Search completed: September 5, 2003, 13:17:57
Job time: 3101 secs